THE UNIVERSITY OF
WARWICK

University of Warwick institutional repository: http://go.warwick.ac.uk/wrap

A Thesis Submitted for the Degree of PhD at the University of Warwick

http://go.warwick.ac.uk/wrap/3084

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it. Our policy information is available from the repository home page.

Expression of Corticotrophin-Releasing Hormone Receptor Subtypes in Human Myometrium and Cloning of the Promoter Region for the CRH Receptors Type 2.

Jing Chen

A Thesis submitted in partial fulfilment of the university's requirements

for the degree of Doctor of Philosophy

September, 2001

Department of Biological Sciences

University of Warwick

Contents

List of tables

List of figures

Acknowledgements

I would like to thank my supervisor, Professor Edward Hillhouse, for all his supervision, encouragement and advice during the process of my experimental work and the preparation of the thesis.

I would like to thank Dr Andrew Easton for his encouragement, helpful discussions and advice. Also, I would like to thank Dr Dimitris Grammatopoulous for his advice and support.

I would like to thank my wife, Jihong and my daughter Zhongbo, for all their love and support.

Declaration

The work described is the result of my own independent research except where specifically acknowledged in the text. All sources have been acknowledged by means of a reference.

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or any other university.

Abbreviations

A – adenine (base) or adenosine (nucleoside)

aa – amino acid (s)

ACTH – adrenocorticotropin hormone

ATP – adenosine triphosphate

dATP – deoxyadenosine triphosphate

bp – base pair (s)

BSA – bovine serum albumin

C – cytosine (base) or cytidine (nucleoside)

Ca^{2+} - calcium ion

CAT – chloramphenicol acetyltransferase

$CaCl_2$ – calcium chloride

cDNA – complementary DNA

Ci – Curie

cAMP – cyclic adenosine 3',5'-monophosphate

CIAP – Calf intestinal alkaline phosphatase

CRH - Corticotrophin releasing hormone

CRH-R - Corticotrophin releasing hormone receptor

CTP – Cytidine triphosphate

dCTP – Deoxycytidine triphosphate

DMEM – Dulbecco's Modified Eagle Medium

DMSO - dimethylsulfoxide

DNA – Deoxyribonucleic acid

DNase – Deoxyribonuclease

DTT – Dithiothreitol

EDTA – ethylenediaminetetraacetate acid

G – guanine (base) or guanidine (nucleoside)

GPCR – G-protein coupled receptor

GTP – Guanosine triphosphate

dGTP – Deoxyguanosine triphosphate

g – gram

G protein – guanine nucleotide-binding protein

HET – Human Kidney transformed

IPTG – Isopropyl-B-D-thiogalactosidase

kb – Kilobase (s) or 1000 bp

LB – Luria-Bertani media

mg – Milligram

$MgCl_2$ – magnesium chloride

ml – Millilitre

MOPS – 3-(N-morpholino) propanesulphonic acid

mRNA – Messenger RNA

N – any nucleotide

ng – Nanogram

NH_4OAc – ammonium acetate

nt – nucleotide (s)

NTP – nucleotide triphosphate

dNTP – deoxynucleotide triphosphate

oCRH – ovine CRH

OD – Optical density

ORF – Open reading frame

PBS – phosphate buffered saline

PCR – polymerase chain reaction

PKA – protein kinase A

PKC – protein kinase C

R – a purie, i.e. A or G

5'RACE – rapid amplification of cDNA 5'ends

r/h CRH – rat/human CRH

RNA – ribonucleic acid

RNase – ribonuclease

rpm – revolutions per minute

RT – reverse transcription

SDS – Sodium dodecylsulphate

SSC – Standard saline citrate

T – thymine (base) or thymidine (nucleoside)

TAE – Tris-acetate-EDTA buffer

TBE – an electrophoresis buffer containing Trs, borate and EDTA

TdT – terminal transferase

TE – Tris-EDTA

TEMED – N,N,N',N'-Tetramethylethylenediamine

TM – Transmembrane

Tris – 2-amino-2-(hydroxymethyl)-1,3-propandiol

dTTP – Deoxythymidine triphosphate

U - Unit

µg – Microgram

µl – Microlitre

5'UTR – 5'untranslated region (s)

UTP – uridine triphosphate

v/v – volume:volume ratio

w/v – weight:volume ratio

X-Gal – 5-bromo-4-chloro-3-indolyl-β-galactopyranoside

Y – a pyrimidine, i.e. C or T

Summary

Corticotrophin releasing hormone (CRH) and CRH receptor (CRH-R) appears to play a number of important roles in human pregnancy. The purpose of the first part of my project was to clone and sequence CRH-R subtypes from human myometrial biopsies. In order to understand the molecular mechanisms that direct the expression of the CRH-R gene, the objective of the second part of my research project was to clone and characterise the promoter region of the human CRH-R2 gene.

Our results demonstrated the presence of multiple CRH-R mRNAs in the human myometrium. Six subtypes of the CRH receptor, 1α, 1β, 1c, 2α, 2β, and 2γ, were found in the pregnant human myometrium, whereas only three subtypes, 1α, 1β, and 2β were found in the nonpregnant myometrium. Multiple CRH-R mRNAs have been identified in human myometrium with differential expression pattern during pregnancy, which argues for multiple roles for CRH and/or related peptides in myometrial function and suggests distinct functional roles for each receptor during pregnancy. These findings suggest that CRH and its receptor play an important modulatory role in myometrial function.

The genomic organisation of human CRH-R2 alternative exons has been determined in the project. The coding region of the gene spans over 18 kb. The genomic orders of the alternative exons are CRH-R2β-2γ-2α. The 5'-flanking region of the human CRH-R2 gene was cloned by the genomic walking method. Using 5'RACE, the transcription start sites were mapped. The results suggest that there may be a single promoter regulating the expression of all (2α, 2β, and 2γ) subtypes. The CRH-R2 5'-flanking sequence has many characteristics of a housekeeping gene promoter. Therefore, we speculated that the strong and housekeeping gene-like activity of he CRH-R2 gene promoter may contribute to the ubiquitous expression of the CRH-R2 gene. For functional analysis, 5'-flanking sequences up to-1393 bp were fused to the Chloramphenicol acetyltransferase (CAT) gene and tested using a HEK-293 cell line by transfection CAT assays. The results confirmed that the DNA region indicated demonstrated strong basal promoter activity.

Chapter 1. General description of corticotrophin-releasing hormone and its receptor

1.1 Introduction

In 1955, experimental support for the hypothesis that adrenocorticotropic hormone (ACTH) secretion is controlled by hypothalamic factors (Harris GW 1948) was obtained by Guillemin and Rosenberg. However, the chemical nature of this factor remained unknown until 1981 when a 41 amino acid C-terminal amidated peptide was purified from ovine hypothalami and shown to be capable of stimulating pituitary ACTH release *in vitro* (Vale *et al.*, 1981). The biologically active form of this peptide, designated corticotrophin-releasing factor-41 (CRF-41), and also frequently referred to as corticotrophin-releasing hormone (CRH), was synthesised and found to have potent ACTH-releasing actions *in vivo* (Vale *et al.*, 1983).

Corticotrophin-releasing hormone, which is termed CRH in this thesis, is a 41 amino acid peptide which plays a major role in co-ordinating the behavioural, autonomic, endocrine and immune responses to stress. CRH mediates its effect via heterotrimeric guanine nucleotide-binding protein (G-protein) coupled receptors which belong to the large superfamily of seven-transmembrane domain receptors and transduce their signal through stimulation of adenylate cyclase and cAMP production (Perrin *et al.*, 1986, Battaglia *et al.*, 1987).

Molecular cloning studies indicate the existence of at least two major types of mammalian CRH receptors (CRH-R), termed CRH-R1 and CRH-R2. CRH-R1 has four subtypes, namely, CRH-R1α, CRH-R1β, CRH-R1c and CRH-R1d. CRH-R1α is a 415-amino acid protein that is expressed mainly in the brain (Vita *et al.*, 1993) and pituitary (Chen *et al.*, 1993). Three spliced variants of CRH-R1α have been identified, which are CRH-R1β (Chen *et al* 1993), 1c (Ross *et al* 1994) and 1d (Grammatopoulos *et al.*, 1999). CRH-R2 (Liaw *et al.*, 1996), which shares 70% identity with CRH-R1, is expressed in the form of three functional splice variants, each of which has alternative 5'exon splicing: CRH-R2α (411 amino acids), 2β (438 amino acids) and 2γ (397 amino acids) (Kishimoto *et al.*, 1995, Lovenberg *et al.*, 1995

a, Stenzel *et* al., 1995, Kostich *et al.*, 1998). Human CRH-R2α and CRH-R2β are co-expressed in peripheral organs and the central nervous system (CNS), whereas CRH-R2γ, which has been isolated only from humans, was found only in the brain (Kostish *et al.*, 1998).

These receptors show differences both at the level of their expression between tissue types and in their pharmacological properties; For example, CRH-R1 binds CRH and CRH-like peptides (Urocortin, Urotensin 1, oCRH, r/hCRH and Sauvagine) with equal affinity, whereas CRH-R2 preferentially binds Urocortin, suggesting that Urocortin is the native ligand. The difference in the pharmacological profile and distribution of CRH-R1 and CRH-R2, and their splice variants suggest that they may have different functions.

It has been shown that CRH-R1 and CRH-R2 are encoded by separate genes. The gene coding for CRH-R1 consists of at least 13 exons and spans over 20 kilobases (Sakai *et al.*, 1998). The CRH-R1 gene has been assigned to 17q12-q22 (Vamvakopoulos and Sioutopoulou, 1994). The gene coding for CRH-R2 consists of at least 12 exons and spans approximately 30 kilobases (Liaw *et al.*, 1996). The gene has been mapped to the chromosome 7p21-p15 (Meyer *et al.*, 1997).

Human CRH receptors (hCRH-Rs) have been identified in many tissues such as the central nervous system (CNS), spleen, ovary, adrenal, and myometrium. The wide distribution of the CRH-Rs suggests that, in addition to its role in regulation of the hypothalamic-pituitary-adrenal (HPA) axis, CRH may play an important role in the physiology of many other organs. Recent studies have implicated CRH and its receptors in the aetiology and pathophysiology of disorders such as depression, anxiety disorders, anorexia nervosa, neurodegenerative diseases and inflammatory illnesses such as rheumatoid arthritis (De Souza EB 1995).

The initiation of parturition in humans remains an enigma, but it seems to be multifactorial. Recently, interest has been directed toward placentally derived corticotrophin-releasing hormone (CRH). Although CRH concentrations in the human peripheral circulation are normally low, they increase throughout pregnancy and fall

rapidly after parturition. The physiological role of the these high concentrations of CRH during gestation is unknown, but it has been suggested that there is a "CRH placental clock" which determines the length of gestation and the timing of parturition and delivery (McLean *et al.*, 1995). Placentally derived CRH influences myometrial function via specific CRH receptor.

Evidence for a direct action of CRH on human myometrium was provided by the identification of a single, high affinity population of CRH receptors in the human myometrium which increase their affinity during the latter stages of pregnancy (Hillhouse *et al.*, 1993) and become functionally linked to the adenylate cyclase system, leading to the increased production of cAMP, a well known myometrial relaxing factor (Grammatopoulos *et al.*, 1994). At term, there is a reduction in the functional ability of CRH to stimulate cAMP due to (1) a reduction of the number of $G_S\alpha$ subunits and (2) an inhibitory action of oxytocin, which activates protein kinase C, leading to phosphorylation and desensitisation of the CRH-R (Grammatopoulos and Hillhouse, 1999).

In particular, it has been demonstrated (Grammatopoulos *et al.*, 1995) that the human myometrium expresses at least five isoforms of the CRH receptor: pI 4.65, 4.8, 4.95, 5.1, and 5.2. These results suggest that the presence of different isoforms may subserve different properties and functions (Grammatopoulos *et al.*, 1995). CRH may play a role by promoting myometrial relaxation during pregnancy and by facilitating myometrial contractions during labour. These effects are likely to be mediated via different receptor subtypes. It is possible that alterations in mRNA processing produce different spliced variant observances.

1.2 Corticotrophin-releasing hormone (CRH)

CRH is a 41 amino acid, single-chain polypeptide, which has now been described in many species (including rat, human, pig, sheep, horse, goat, cow, Xenopus, and white suckerfish). CRH has sequence homology and shares many biological properties with two other vertebrate peptides: sauvagine, a 40 amino acid peptide isolated from the

frog skin (Montecucchi *et* al, 1981), and urotensin I, a 41 amino acid peptide isolated from the caudal neurosecretory system of teleost fish (Lederis *et al.*, 1982). CRH also has sequence homology to two diuretic hormones, Mas-DPI and MAS-DPII, from the tosacco hornworm, *Manduca sexta* (Katoaka *et al.,* 1989, Blackburn *et al.,* 1991). The vertebrate homologies have been tested and found to possess potent mammalian and fish pituitary ACTH-releasing activity. In addition, they decrease peripheral vascular resistance and cause hypotension when injected into mammals (Rivier *et al.,* 1983, Lenz *et al.*, 1985). CRH is produced and secreted primarily from parvocellular neurones of the paraventricular nucleus (PVN) of the hypothalamus. It is abundant in PVN nerve terminals within the median eminence, where it is released into the hypophyseal portal circulation.

CRH plays a major role in co-ordinating the behavioural, autonomic, endocrine and hormonal responses to stress. The major endocrine response to stress is via activation of the hypothalamic-pituitary-adrenal (HPA) axis, leading to rapid secretion of adrenocorticotropin (ACTH) by the pituitary corticotrophin and increases in circulating glucocorticoids, which are critical for adaptation (Figure 1.1). When the same stressor (homotypical) is repeated, the pattern of activation differs according to the type of stress, with desensitisation of ACTH responses to repeated homotypical stimuli for some stressors (such as cold exposure and repeated immobilization), or preservation of the responses after each homotypical stimulus for others (such as repeated foot shock, insulin hypoglycemia and intraperitoneal hypertonic saline injection) (Aguilera, G. 1998). In addition to its endocrine effects, immunohistochemical localisation of CRH has demonstrated that the hormone has a broad extrahypothalamic distribution in the central nervous system (CNS) (Olschowka *et al.,* 1982, Cummings *et al.,* 1983). These effects of CRH are mediated by adenylate cyclase/cAMP/protein kinase A-dependent mechanisms triggered by binding of CRH to CRH receptors (CRH-R) in the corticotrophin cell (Abou-Samera *et al.*, 1987).

CRH produces a wide spectrum of autonomic, electrophysiological, and behavioural effects which are consistent with a neurotransmitter or neuromodulator role in the brain. Intracerebroventricular administration of CRH provokes stress-like responses, including activation of the sympathetic nervous system and inhibition of the

parasympathetic nervous system (Brown *et al.*, 1982). The behavioural profile following central administration of CRH is also characteristic of a compound which increases arousal and emotional reactivity to the environment. These effects of CRH include general arousal, as exhibited by increased locomotion, sniffing, grooming, and rearing in familiar surroundings and increased agitation in unfamiliar surroundings (Dunn AJ and Berridge CW 1990). Recent evidence suggests that CRH may also play a significant role in a variety peripheral functions, including cardiovascular function, inflammation, reproduction, and integration of the immune system's response to stressors (Blalock *et al.*, 1989).

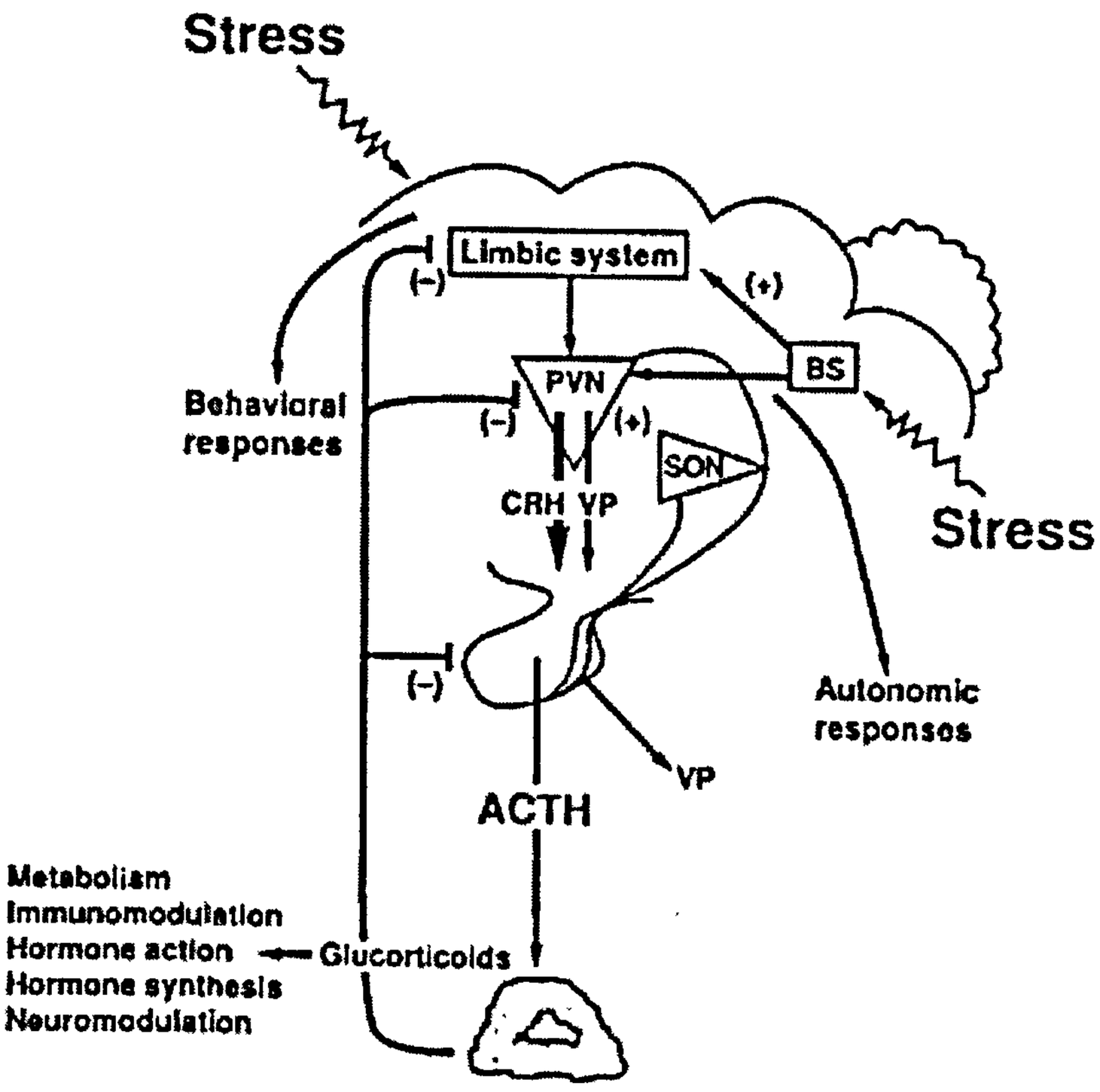


Figure1.1. Diagram illustrating the integrated responses to stress emphasising the activation of the hypothalamic-pituitary-adrenal (HPA) axis and the feedback and regulatory effects of glucocorticoids. Activation of the supraoptic nucleus (SON) and magnocellular paraventricular nucleus (PVN) with release of vasopressin (VP) from the neurohypophysis occurs only during osmotic stress. ACTH, adrenocorticotropin; BS, brain stem; CRH, corticotrophin-releasing hormone (Aguilera, G. 1998).

1.3 Corticotrophin releasing hormone receptor (CRH-R)

1.3.1 CRH-R subtypes

The function of CRH is mediated via its specific interactions with cell surface receptors. Molecular cloning studies indicate the existence of at least two major types of mammalian CRH-R, (CRH-R1 and CRH-R2). Recently, cDNAs corresponding to rat, mouse, and human CRH-Rs have been characterised. The CRH-Rs belong to the recently described family of "gut-brain" neuropeptide receptors and exhibit significant homology to the parathyroid and glucagon receptor family that has been designated as members of Class II in the G-protein coupled seven transmembrane helical domain receptors (GPCR) superfamily. The CRH-Rs are GPCRs and have been shown to be coupled to both adenylate cyclase and phospholipase C, stimulating cAMP production and phosphoinositide hydrolysis (Segre, G. V. and Goldring, S. R. 1993).

1.3.1.1.CRH-R1

The CRH-R1α has been cloned from several species including human (Chen *et al.*, 1993), mouse (Vita *et al.*, 1993), chicken (Yu *et al.*, 1996) and rat (Perrin *et al.*, 1993, Chang *et al.*, 1993). Species homologues are 98% identical over their full length of 415 amino acids. In general, the CRH-R1 shares approximately 30% identity with all other members of the neuropeptide receptor family. The structure of the cDNA encoding the human pituitary CRH-R1α has been characterised using expression cloning of cDNA from a human Cushing's corticotrophin adenoma (Chen *et al.*, 1993). The CRH-R1α is a 415 amino acid protein comprising seven putative membrane-spanning domains characteristic of most G protein-coupled receptors. The CRH-R1α has putative N-linked glycosylation sites in the N-terminal extracellular domain. There are five such sites predicted, substantiating the glycosylation profiles determined by chemical affinity cross-linking studies (Grigoriadis and De Souza, 1989). In addition, there are potential protein kinase C phosphorylation sites in the first and second intracellular loop and in the C-terminal tail, as well as casein kinase II

and protein kinase A phosphorylation sites in the third intracellular loop (Chen *et al.*, 1993).

Three variant human CRH-R1 cDNA clones have been identified from a corticotrophin adenoma (Chen *et al.*, 1993), myometrium (Grammatopoulos *et al.*, 1999) and hippocampus (Ross *et al.*, 1994). The former encodes 29 additional amino acids inserted in the first intracellular loop (named: CRH-R1β), and the latter has a 40 amino acid deletion in the N-terminal extracellular domain (CRH-R1c). The amino acid sequence of CRH-R1d is identical to the CRH-R1α receptor except that it contains an exon deletion resulting in the absence of 14 amino acids in the predicted seventh transmembrane domain (Figure1.2). Interestingly, studies in which COS-7 cells have been transfected with CRH-R cDNAs have shown that the CRH-R1α is coupled efficiently to both adenylate cyclase and phosphoinositol hydrolysis whereas the spliced variant receptor (CRH-R1β) is not well-coupled to adenylate cyclase and does not couple at all to phospholipase C (Xiong *et* al., 1995). The 29 amino acid insertion strongly reduces the coupling efficiency of the produced protein. Since the only difference between the CRH-R1α and CRH-R1β is the insert in the first putative cytoplasmic loop, these data indicate that the first cytoplasmic loop plays a crucial role in CRH-R1 coupling to the G protein. CRH-R1c failed to bind to [^{125}I] CRH with high affinity in transfected COS-7 cell. This deletion results in a low affinity CRH -R protein which is only activated at high (>300 nM CRH) agonist concentrates.

Recently, Hillhouse and his team have isolated and characterised a novel CRH-R (CRH-R1d) spliced variant from human myometrium by employing a RT-PCR technique. This variant of the CRH-R1α contains a 42 bp deletion in the last transmembrane domain. In HEK-293 cells stably transfected with this variant both CRH and urotensin, but not sauvagine or urocortin, can elevate intracellular cAMP levels at relatively high concentration. In contrast to CRH, urotensin and sauvagine can elevate high levels of intracellular cAMP in HEK-293 cells stably tranfected with CRH-1α (Grammatopoulos, *et al.*, 1999).

In contrast to the human, these splice products have not been identified in rodents. In rats, a CRH-R1 frame shift mutant encoding a 224 amino acid protein has been

identified (Chang *et al.*, 1993). This mutant lacks the entire third and parts of the fourth putative transmembrane domain. At present, nothing is known about the functional properties of this mutant receptor protein.

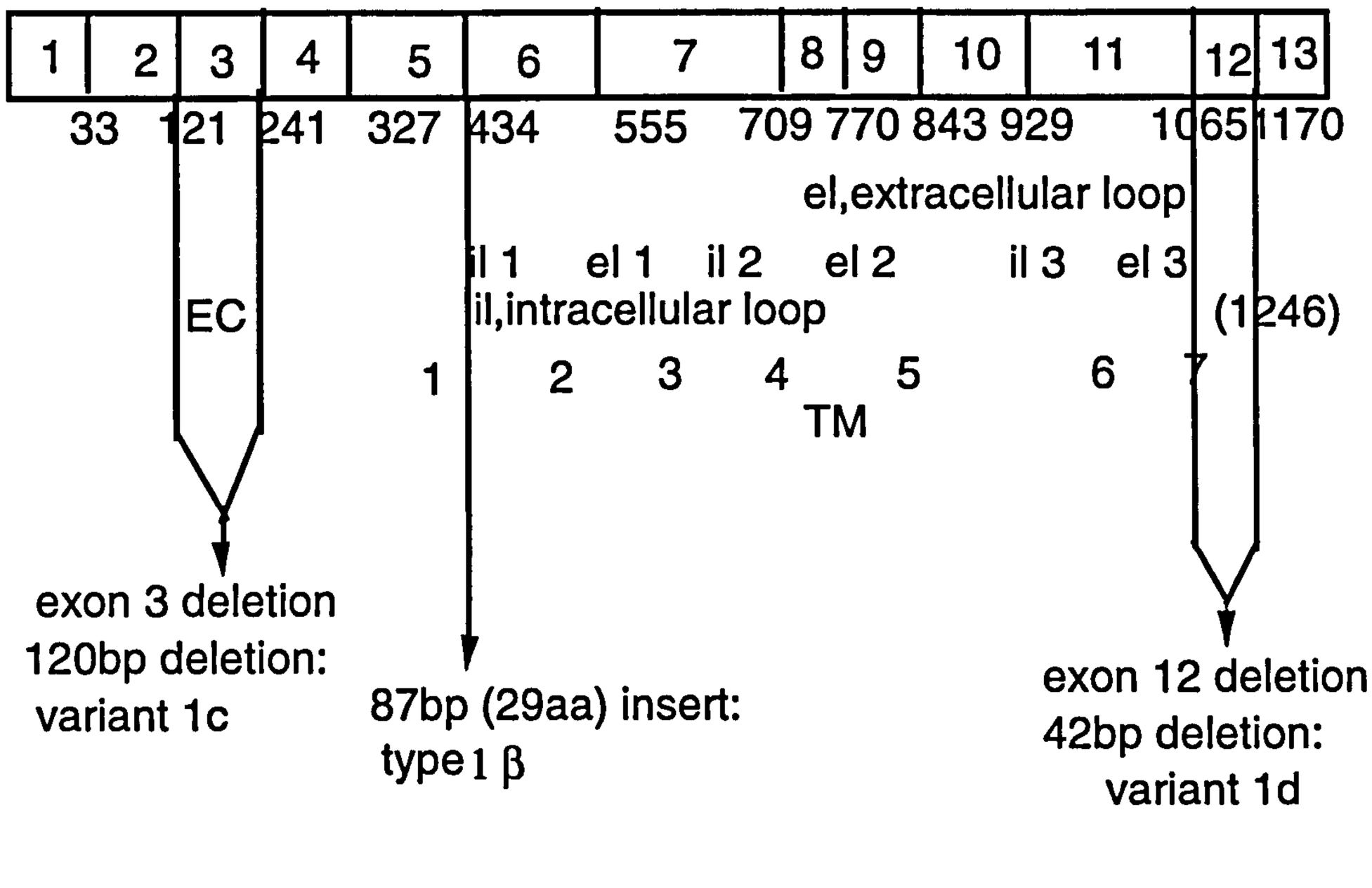


Figure 1.2. Genomic structure of human CRH-R1

The diagram shows the position of CRH-R1 introns in relation to extracellular and transmembrane domains, where EC, extracellular; il, intracellular loop; transmembrane (TM) domains; el, extracellular loop cytoplasmic domain after TMVII (1-1246 bp): initiation to terminating codons. The numbers under the boxes refer to nucleotide sequence numbers (hCRH-R1α: GenBank Accession No.L23332: Chen *et al.*, 1993). The gene coding for CRH-R1 consists of at least 14 exons and spans over 20 kilobases (Sakai *et al.*, 1998).

1.3.1.2 CRH-R2

A second molecularly distinct CRH-R has been identified in rat (Lovenberg *et al.*, 1995a), mouse (Perrin *et al.*, 1995, Stenzel *et al.*, 1995, Kishimoto *et al.*, 1995), and human (Liaw *et al.*, 1996). In the rat at least, there are two known forms of the second

CRH-R. CRH-R2α is a 411 amino acid protein, which is expressed almost exclusively in brain. The alternatively spliced form, CRH-R2β, has the first 34 amino acids of R2α replaced with a 54 amino acid sequence, giving a full-length receptor of 431 amino acids. CRH-R2β shows the highest level of expression in the heart and skeletal muscle and a lesser extent is expressed in brain, lung, and intestine (Lovenberg *et al.*, 1995b). In the mouse, only the homologue of the rat CRH-R2β form has been identified (Perrin *et al.*, 1995, Stenzel *et al.*, 1995, Kishimoto *et al.*, 1995). In the human, the high homology of the reported rat and mouse CRH-R2β form has been described in Valdenaire *et al.*, 1997 (Figure 1.3).

CRH-R2β

h:MRGPSGPPGLLYVPHLLLCLLCLLPPPLQYAACQSQMPKDQPLWALLEQYCHTIMTLTNLSGPYSYCNTTL

r: M- H- --S- PSAQ- - - - - YS- - -++-- -LQV- -QPGRPLQ- - - - - - T- - - - - - - - RT+ --TR- -F- - - - - - - - - - - -

m: M- T- - S- PSAQ- - - - - FS- - - ++--VLQV- -QPG- A- Q- - - - - -T- - - - - - - - RT+ --TG- -F- - - -T- - - - - --

Figure 1.3. Alignment of the amino acid sequences of the 5'end of the human (h), rat (r), and mouse (m) CRH-R2β. A dash (-) represents a residue conserved in the CRH-R2β sequence, while a cross (+) indicates a deletion.

The three splice variants diverge at the corresponding position of human CRH-R2 where the first intron occurs. CRH-R 2 is expressed in the form of three functional splice variants, 2α (411 amino acids), 2β (438 amino acids), and 2γ (397 amino acids) (Liaw *et al.*, 1996, Valdenaire *et al.*, 1997, Kostich *et al.*, 1998). Structural comparison of these CRH-R2 subtypes showed that 377 amino acids at their C terminus are identical and they differ only their N terminus; the 34 amino acids N terminal to CRH-R2α are replaced by a 61 amino acid sequence to form the CRH-R2β or a 20 amino acid sequence to form the CRH-R2γ. CRH-R2α and 2β are co-expressed in peripheral organs and the CNS (Valdenaire *et al.*, 1997), whereas CRH-R2γ, which has been isolated only from humans, was found only in brains (Kostich *et al.*, 1998). Interestingly, only CRH-R2α was isolated from amphibian species (Dautzenberg *et al.*, 1997), whereas in rodents an additional CRH-R2β form has been detected (Lovenberg *et al.*, 1995a).

In adenylate cyclase activation assays, CRH-related peptides such as sauvagine and urocotin are 10-fold more potent at the CRH-R2β than CRH-R2α or CRH-R2γ, which suggests that the N-terminus of the receptor is involved in the ligand receptor interaction (Kostich *et al.*, 1998). In these experiments, urocortin was found to be the most potent peptide for activation of the CRH-R2.

CRH-R2 contains similar putative extracellular N-glycosylation and transmembrane phosphorylation sites to CRH-R1. CRH-R1 and CRH-R2 are 70% identical, with greater sequence homology toward the C-terminus and only 47% identity in the N-terminal extracellular domain (Liaw *et al.*, 1996). It is interesting to note that there are very large regions of amino acid identity between CRH-R1 and CRH-R2, particularly between transmembrane domains five and six. This similarity underlies the conservation of second messenger function, since it is this region that is thought to be the primary site of G protein coupling.

Very recently, Miyata *et al* (1999) have identified and characterised a cDNA encoding a novel isoform of the CRH-R, referred to as CRH-R2α-tr, from a rat amygdala cDNA library. The nucleotide sequence of the cloned cDNA has a structure of an alternatively spliced form of the CRH-R2α, which contains unspliced introns 6 and 7 in the message, and encodes a 236 amino acid truncated protein that comprises three unique transmembrane domains (Miyata *et al.*, 1999). In contrast to rodents, this splice product has not been identified in humans.

1.3.2 Seven transmembrane helical domain receptor

Hormones can be generally divided into three major classes: (1) derivatives of amino acids (e.g., dopamine and catecholamines), (2) peptides and protein (e.g., CRH), and (3) derivatives of steroids (e.g., estrogen). For the most part, amino acid derivatives and peptide hormones interact with membrane receptors on the cell surface, whereas the steroid hormones act by crossing the plasma membrane to interact with intracellular receptors.

Membrane receptors: receptors on the surfaces of cells provide an essential component of communication with extracellular ligands. These membrane receptors

fall into three general classes based on their structure and mechanisms of signal transduction: (1) the single transmembrane domain receptors; (2) the four transmembrane domain receptors; (3) the seven transmembrane domain (7TM) receptors. The letter receptors have a conserved structure that is characterised by seven hydrophobic domains that are inserted into the membrane (Figure 1.4).

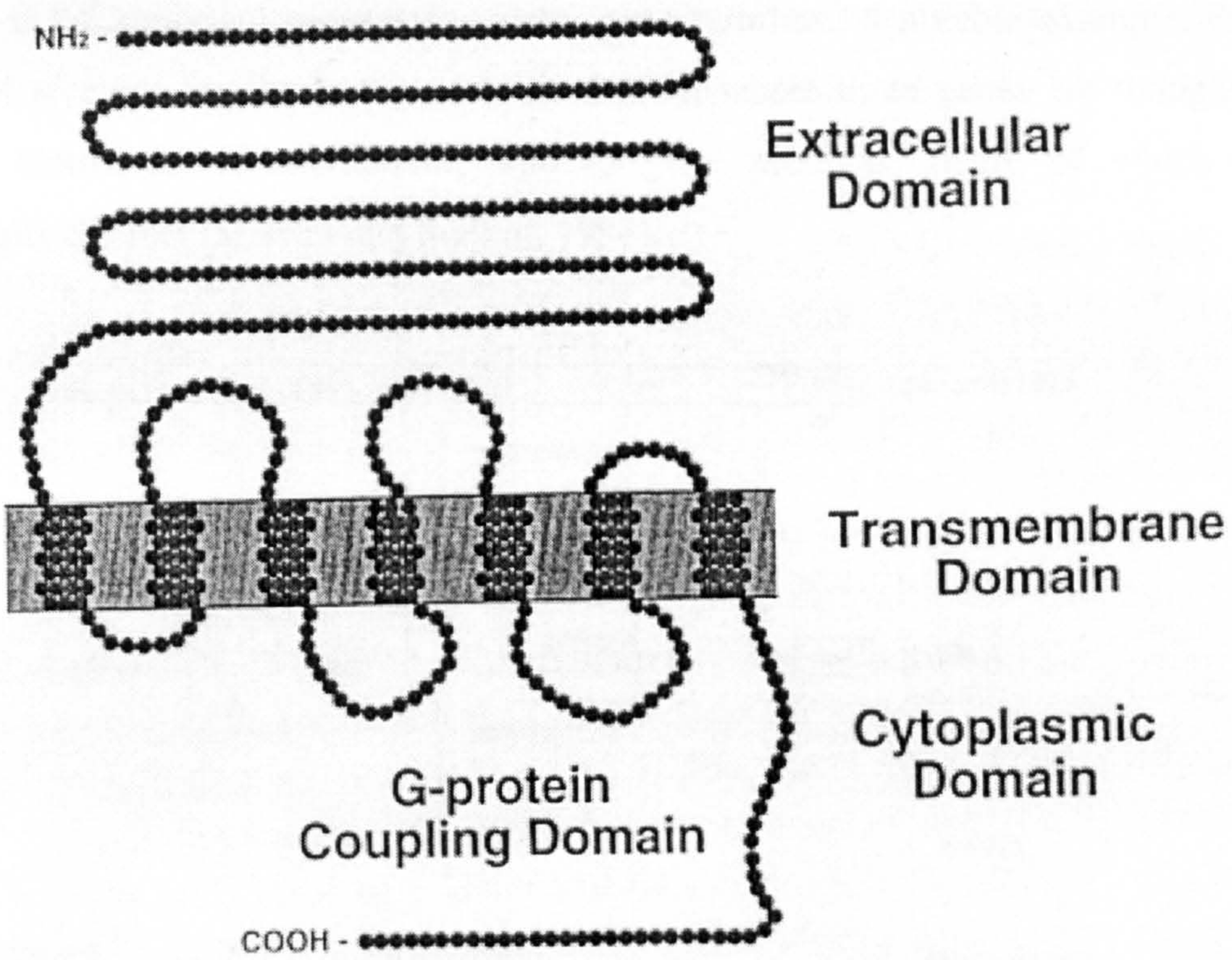


Figure 1.4. Structure of a seven transmembrane domain receptor.
The G protein coupled receptors share several structural features, including the characteristic seven transmembrane spinning domains. The amino-terminal (NH_2) domain is located extracellularly and is involved in hormone interaction for some receptors. The seven transmembrane domains are connected by intra- and extracellular loops of amino acids. In many receptors, the ligands bind within the transmembrane domains. The intracellular loops and carboxy-terminal tail (COOH) are involved in coupling to G proteins and in intracellular signalling.

Seven transmembrane domain receptors exhibit large diversity within the family and several 7TM receptors may recognise the same endogenous ligand. For example,

eight unique 7TM receptors have been identified for glutamate. Many 7TM receptor genes contain multiple introns, which introduce the possibility of variants due to alternative splicing, exon skipping and intron retention. Variations in gene-product splicing occur at the level of processing of heteronuclear-RNAs (hn RNAs or pre-mRNAs) in the cell nucleus. In genes with multiple exons, the introns are removed. Alternative splicing pathways vary the particular introns to be deleted (Figure 1.5) and this is an important process for many gene families. A notable example, for a non-7TM receptor family, is the neurexin family, where three genes are thought to produce thousands of alternatively spliced gene products, many of which are functionally distinct (Missler and Sudhof, 1998).

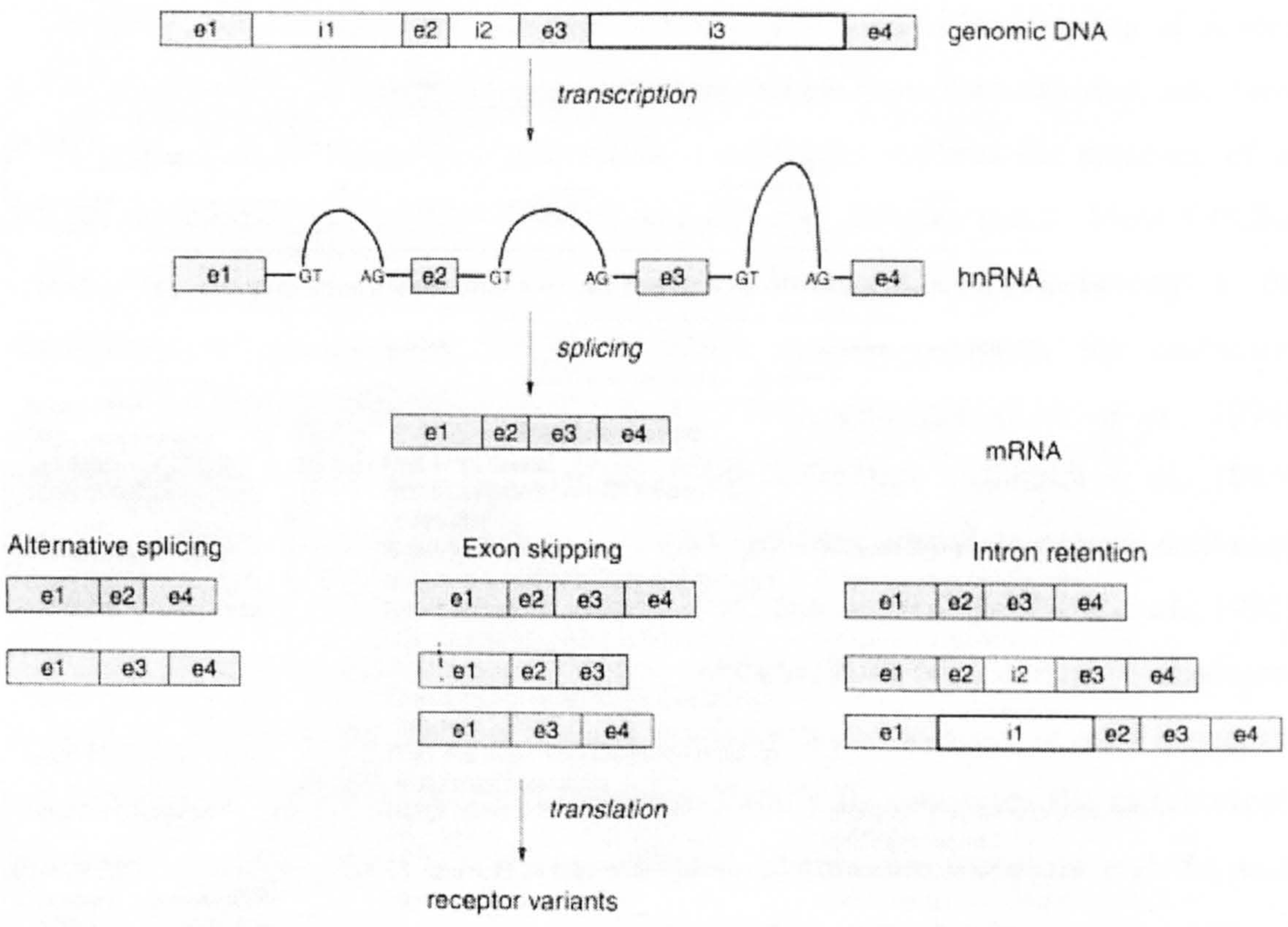


Figure 1.5 The process by which splice variants can be produced.

e1-e4, exons 1-4 of a hypothetical gene; i1-i3, introns 1-3 of a hypothetical gene. Gene transcription results in synthesis of a single stranded linear unit of RNA, called hnRNA, which contains all of the sequence encoding the exons and introns. Introna are removed and the remaining exons joined together at defined base sequences by a process called splicing. Many 7TM receptor genes

contain multiple introns, which introduce the possibility of variants due to alternative splicing, exon skipping and intron retention (Kilpatrick *et al.*, 1999).

It is possible that 7TM splice variants could allow further regulation of receptor function via different dimerization possibilities and additional regulation of distribution within cell via differential interaction with intracellular scaffolding proteins

1.3.3 G proteins and G-protein-coupled receptors

Molecular cloning approaches have led to the isolation and identification of several hundred individual G-protein-coupled receptors, which have been divided, into three subfamilies. The inclusion of a receptor in a subfamily requires the presence of an overall percentage amino acid identity and not any discrete motif. Most GPCRs, including the odorant receptors, are grouped in Family I. CRH-Rs belongs to the superfamily of neuropeptide receptors, which include receptors for vasoactive intestitial peptide (VIP) (Sreedharan *et al.*, 1995), glucagon (Lok *et al.*, 1994), parathyroid hormone (PTH) (Kong *et al.*, 1994), calcitonin (Albrandt *et al.*, 1995), growth hormone releasing factor (Mayo, 1992), pituitary adenylate cyclase-activating polypeptide (PACAP) (Arimura and shioda 1995) and secretin (Ishihara *et al.*, 1991). The CRH-R1 and R2 complementary DNAs (cDNAs) have been cloned (as discussed in Section 1.3.1) and the encoded protein has been found to share several conserved features with other GPCR family II members. Family III comprises the metabotropic glutamate receptors, the Ca^{2+}-sensing receptor, pheromone receptors and the taste receptors. Within each family, GPCRs are grouped by sequence similarity and ligand specificity.

The GPCRs are integral membrane proteins involved in the transmission of signals across membranes, When a ligand binds the GPCRs undergo one or more conformational changes that trigger interactions between the receptor and G proteins. These initiate the intracellular signalling response and modulate the activity of distinct effector systems such as adenylyl cyclase, phospholipases, cGMP, phosphodiesterases

and ion channels. GPCRs also undergo desensitization, becoming refractory to further stimulation after the initial response despite the continued presence of a stimulus of constant intensity. Phosphorylation is a critical event in regulating this process (Kaziro *et al.*, 1991).

G proteins are heterotrimers that consist of an α-subunit and a combined β-γ-subunit. There are at least twenty three different α-subunit genes as well as several different β (6β)- and γ (12γ)-subunits. Within any given cell type, multiple different receptors, G protein, and effectors are present. This diversity of G protein subunits requires specific interactions that appear to be mediated by the intracellular loops and C-terminus of the receptor. In the absence of ligand, the α-subunit is bound to guanosine diphosphate (GDP) and remains associated with the β-γ-subunits. Ligand binding to the receptor induces a conformational change in the receptor that promotes the binding of a G protein to the receptor. After coupling to the receptor, GDP is exchanged for GTP on the α-subunit and the β-γ-subunit dissociates. These steps activate the G protein and initiate interactions with downstream effector molecules (e.g., adenylate cyclase). The α-subunit contains an intrinsic GTPase activity that converts GTP to GDP, which restores the inactive state and allows reassociation with the β-γ-subunits.

Recent studies have indicated that alternative splicing is responsible for the generation of a number of G protein coupled receptor isoforms. The Human Parathyroid Hormone (PTH)/PTH-related peptide (PTHrP) receptor, which is an example of GPCRs, can activate multiple second messenger systems through receptor activation of different G proteins. The multi-exonic structure of the PTH/PTHrP receptor gene raised the possibility that novel PTH/PTHrP receptor variants may arise by alternative splicing of this complex gene (Kong *et al.*, 1994). Several alternatively spliced transcripts may occur in different regions of this gene. The PTH/PTHrP receptor is known to be capable of signalling in response to PTH or PTHrP via two G protein-coupled pathways: (1) Gq-mediated activation of phospholipase C (PLC), resulting in increased Ca^{2+} and activation of protein kinase C (PKC); and (2) Gs-mediated activation of adenylyl cyclase leading to cAMP production and protein kinase A (PKA) activation (Juppner *et al.*, 1991).

Many structure-activity studies with 7TM G-protein-coupled receptors suggest that the concerted participation of several intracellular domains is involved in coupling to G proteins (Spiegel *et al.*, 1992). Findlay *et al* has shown that the C-terminal cytoplasmic tail of the porcine calcitonin receptor is important for efficient coupling to cAMP- and calcium-mediated pathways. Mutations in the third intracellular loop can lead to constitutive activation of receptors (Samama *et al.*, 1993). Differentially spliced PACAP receptors, containing different sequences in this region, show specific coupling to different G proteins (Spengler *et al.*,1993). In addition, a mutation in the putative second intracellular domain of the vasopressin type 2 receptor results in an inability of this receptor to stimulate adenylate cyclase (Rosenthal *et al.*,1993).

1.3.4 Adenylate cyclase

Adenylate cyclase, the enzyme that produces cyclic AMP (cAMP), is widely distributed in mammalian cells and most organisms, and is usually associated with the plasma membrane. Activities of adenylate cyclase are regulated by numerous hormones and transmitters which act through G-protein couple receptors to stimulate or inhibit enzyme activity.

cAMP signalling cascades are typically initiated by the binding of peptide hormones to their membrane receptors. However, intracellular cAMP concentrations are also modulated by intrinsic processes such as the cell cycle. The pleiotropic array of cellular effects of cAMP is primarily mediated by protein kinase A, which acts on a number of cellular substrates, including enzymes, the cytoskeleton and transcription factors. As depicted in Figure 1.6, cAMP binds to a regulatory subunit of protein kinase A, leading to dissociation of an active catalytic subunit that phosphorylates specific substrates.

In addition a group of transcription factors are important targets for protein kinase A pathway (see Section 1.7). These include the cAMP response element binding proteins (CREBs) and activating transcription factors (ATFs) that are members of the β-Zip class of transcription factors. Post-translational modification of CREB by

phosphorylation at serine residues induces conformational changes that alter the affinity of CREB for coactivator proteins, such as CREB-binding protein (CBP) or TATA box-binding protein coactivator $TAF_{II}110$. CBP is thought to form a bridge between CREB and the basal transcription apparatus. CBP also interacts with other β-Zip proteins, such as c-Jun and c-Fos, other transcription factors, including c-Myb, as well as specific kinases (Lowe *et al.,* 1998).

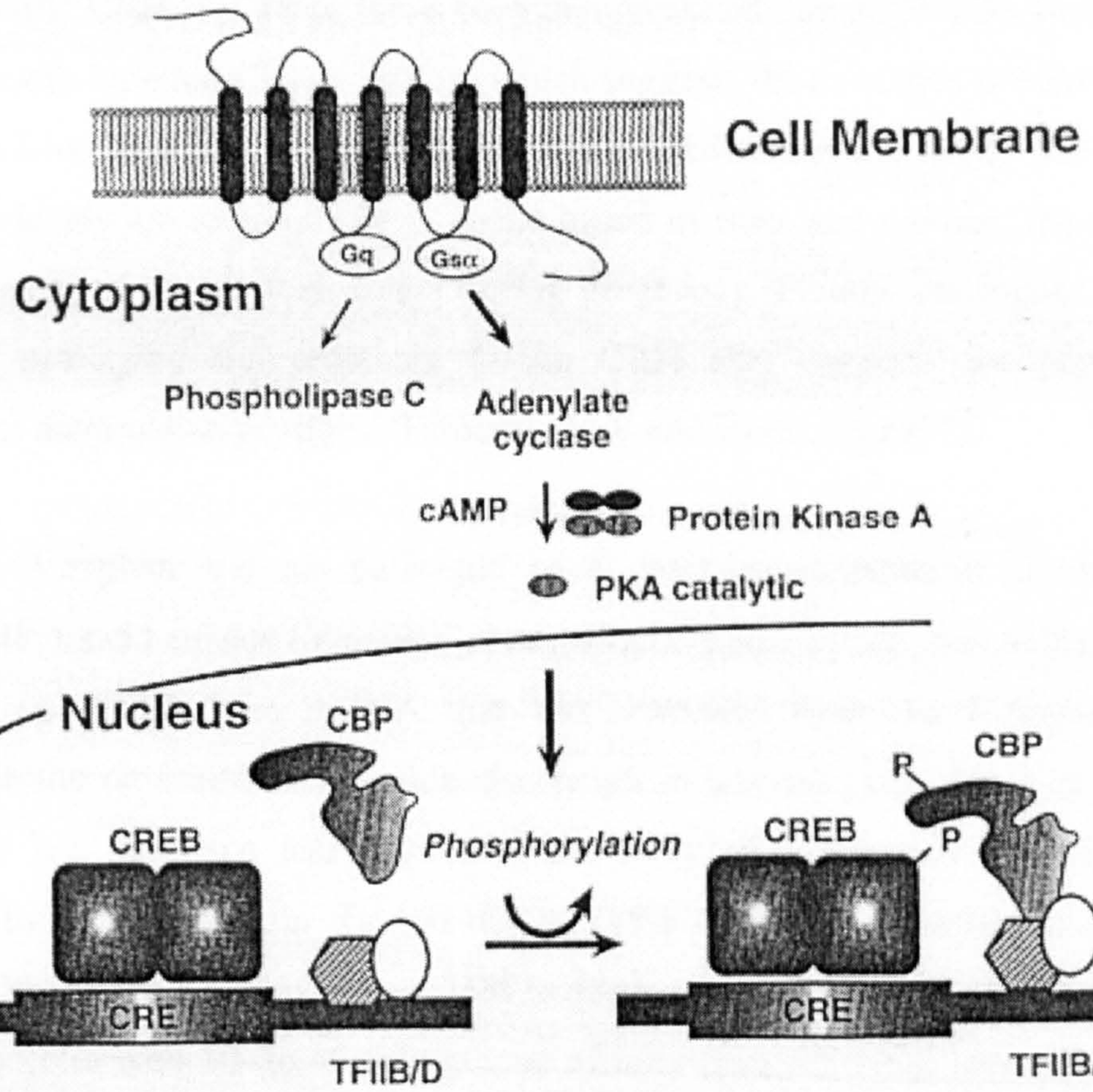


Figure 1.6. cAMP stimulation of transcriptional responses.

G protein coupled receptors provide a major pathway for the generation of cAMP by adenylate cyclase. Increased cAMP levels within the cell cause dissociation of the tetrameric protein kinase A (PKA) holoenzyme leading to the release of the active catalytic subunit of PKA. The active catalytic subunit is translocated to the nucleus, where it phosphorylation transcription factor CREB, which binds as a homodimer to specific cAMP response elements (CREs) on target genes. The phosphorylation of CREB induces the binding of the coactivator, CREB-binding protein (CBP), which is also a substrate for

phosphorylation. These transcription factors act by interacting with basal transcription factors such as TFIIB and TFIID (Lowe *et al.*, 1998).

1.3.5 Alternative CRH ligands

A number of lines of evidence indicate the likely presence of one or more mammalian peptides that are ligands for CRH-Rs / or CRH-BP but are molecularly distinct from hypothalamic $CRH_{(1-41)}$. First, it has been demonstrated that amphibian sauvagine and fish urotensin I are not CRH orthologs which suggests the existence of sauvagine- and urotensin I-like peptides in mammalian species (Okawara *et al.*, 1988). Second, while CRH-BP levels are relatively high in the blood of men and women, the circulating CRH concentrations are low except during pregnancy. Finally, the higher affinity of CRH-R2 sauvagine and urotensin I than CRH also suggests the possibility of equivalent mammalian peptides (Turnbull, A. V and River, C. 1997).

Recently, Vaughan and his colleague have described a putative alternate CRH-R/CRH-BP ligand in rats (Vaughan *et al.*, 1995). This peptide was called urocortin and was identified from mRNA that was extracted from the Edinger-Westphal nucleus in the rat brain; this peptide also exists in humans (Donaldson *et al.*, 1996). Urocortin has sequence and functional similarity to urotensin I (63% sequence identity to carp urotensin I) and CRH (45% sequence identity to r/h CRH) (Figure1.7). Urocortin also shares 35% sequence identity with sauvagine. CRH-R2 binds urocortin with 10- to 40-fold greater affinity than CRH and urocortin has a 10-fold lower EC_{50} than CRH in activating adenylate cyclase in cells overexpressing the CRH-R2β (Vaughan *et al.*, 1995). Indeed, it has been proposed that urocortin is an endogenous ligand for the second CRH-R. Urocortin is also a potent agonist for the CRH-R1 and binds to CRH-BP. The binding coefficient of this interaction is about 2-fold lower than CRH, and urocortin displaces CRH bound to CRH-BP thus elevating "free" CRH levels in brain (Behan *et al.*, 1996).

The comparison of sequences for urocortin, carp urotensin I, CRH, and sauvagine is shown in Figure 1.7.

	1 5 10 15 20 25 30 35 40	
rUcn	DDPPLSIDLTFHLLRTLLELARTQSQRERAEQNRIIFDSV	100%
hUcn	-N--S------------------------------------	95%
cUro	N----I---------- --NMI- M--NEN---Q-GL---KYL-E-	63%
r/hCRH	SEE-- I- L--- -----EV--M--AEQLAQQ-HS---KLMEII	45%
Svg	EG-- I----SLE----KMI- IEKQEKEKQQ-AN---LLL- TI	35%

Figure 1.7 Aligment of the amino acid sequences of the CRH-like peptides rat (r) and human (h) urocortin (Ucn), carp urotensin I (cUro), rat/human CRH, and sauvagine (Svg). -, identical amino acid to rat urocortin; %, percentage sequence homology with rat urocortin.

1.3.6 CRH-binding protein

In the circulation CRH is bound with high affinity by a CRH-binding protein, CRH-BP, which inhibits the biological activity of CRH in a number of *in vitro* assay systems. CRH-BP was originally identified, isolated, and purified from human plasma, and is thought to regulate the biological activity of placental CRH in pregnant woman (Campbell *et al.*, 1987). The human CRH-BP gene has been assigned to 5q11.2-q13.3. Human and rat CRH-BP cDNA display a high degree of sequence homology (85%) and encode a precursor protein of 322 amino acids in length. Both proteins possess one putative N-linked glycosylation site and 11 conserved cysteine residues (Potter *et al.*, 1991). Recently, highly homologous cDNAs encoding mouse and sheep CRH-BP have also been cloned. Although forms of the human and rat CRH-BP are homologous, the pattern of anatomical distribution is somewhat different in the two species. The human CRH-BP is found abundantly in tissues such as the liver, placenta and brain, while in the rat CRH-BP mRNA is found only in the brain and pituitary gland (Potter *et al.*, 1991). The absence of CRH-BP in nonprimate blood presumably reflects the lack of its synthesis in liver. CRH-BP in brain appears to be membrane associated, as indicated by its solubilization profiles and its partitioning upon sucrose gradient purification (Behan *et al.*, 1993). However, there is no evidence for a receptor function for CRH-BP since, unlike the identified CRH-Rs, it lacks obvious transmembrane domains or intracellular signalling motifs in its sequence.

The exact function of CRH-BP under normal or pathological conditions remains uncertain. The pharmacological profile of the CRH-BP is different from that of either the CRH-R1 or CRH-R2. Both human and rat CRH-BP exhibit high affinity for rat and human CRH and very low affinity for ovine CRH (Chalmers *et al.*, 1996). CRH-BP binds urotensin I with extremely high affinity and sauvagine with much lower affinity. Recently, CRH-BP in brain was shown to modulate the availability of "free" CRH to interact with its receptors (Turnbull *et al.*, 1997).

1.3.7 Pharmacology of CRH receptors and CRH-binding protein

The discrete distributions of CRH-R1 and CRH-R2 imply that they subserve distinct physiological functions and this suggestion is supported by the dissimilar pharmacologies of the CRH-R subtypes. Several studies have described the binding constants and potencies of various CRH-R ligands in cells transfected with the cDNAs of the different CRH-Rs. CRH-R1 indiscriminately binds CRH peptides such as urotensin I, rat/human CRH, urocortin and sauvagine with high affinity (Vaughan *et al.* 1995, Donaldson *et al.* 1996). The majority of studies have described sauvagine, urotensin I, and r/h CRH as having similar binding constants or potencies at CRH-R1. CRH-R2 exhibits a substrate specificity that is different from CRH-R1. CRH peptides, such as rat/human CRH or ovine CRH and to a lesser extent urotensin I, are bound with lower affinity than urocortin and sauvagine; the rank order being urocortin > sauvagine > urotensin I > r/hCRH (Kishimoto *et al.*, 1995). CRH-R2α and R2β display virtually identical pharmacological properlies with respect to these ligands. The EC_{50} values derived for CRH-R2α and CRH-R2γ are very similar, whilst those for CRH-R2β are approximately 10-fold lower for all peptides (Kostich *et al.*, 1998). The CRH-R antagonist α-helical $CRH_{(9-41)}$ is a far more effective antagonist at CRH-R2 than at CRH-R1 (Kishimoto *et al.*, 1995). Thus, the pharmacology and distributions of the molecular subtypes of CRH-Rs appear to bear out observations made *in vivo* regarding the pharmacology of CRH responses. In particular, the data discussed above strongly suggest that CRH-R1 is the major receptor isoform regulating pituitary ACTH secretion (Owens *et al.*, 1995). On the other hand, anatomical distribution of the receptors imply that CRH's effects on the vasculature (e. g., vasodilatation, hypotension, and inhibition of injury-induced edema) may be

mediated by CRH-R2. Furthermore, the relative potencies of sauvagine, urotensin I, and CRH are suggestive of CRH-R2 mediated events in several other instances. For example, a rank order of sauvagine > urotensin > CRH is apparent in the induction of hypertension and elevation in plasma catecholamines after injection of peptides or peripheral peptide administration. (Brown *et al.*, 1982).

The pharmacological profile of the CRH-binding protein is different from that of the receptor. Using the ligand immunoradiometric assay (LIRMA), the purified recombinant hCRH-BP was found to exhibit the following rank order of potencies: r/hCRH > alpha-helical ovine $CRH_{(9-41)}$ > $hCRH_{(6-33)}$ > $hCRH_{(9-33)}$ > d-phe $hCRH_{(12-41)}$ > oCRH (Sutton *et al.*, 1995). The rat and human CRH fragments $CRH_{1-41-OH}$, CRH_{6-33} and CRH_{9-33} are inactive at CRH-Rs but have high affinity for the CRH-binding protein. Therefore, although there may be some similarities in the binding domain of the CRH-BP and the CRH-R, they are distinct proteins each with unique and discernible pharmacological characteristics. This ability of the CRH-BP to bind and functionally inactivate CRH is somewhat analogous to the role of high-affinity monoamine-transporter molecules, and represents an additional CRH "modulatory" molecule (Chalmers *et al.*, 1996).

1.3.8 Localisation of CRH receptor subtypes

The availability of nucleotide sequences for CRH-R1 and CRH-R2 has allowed a detailed examination of the regional and cellular distribution of CRH-R subtype mRNA expression using in situ hybridisation histochemistry. The results of these studies indicate heterogeneous distribution patterns of CRH-R subtypes in the brain and peripheral tissues and suggest specific physiological roles for these sites in CRH-related functions.

Although CRH-R1 mRNA expression is very high in neocortical, cerebellar and sensory relay structures, expression of CRH-R2 mRNA is generally confined to subcortical structures. Particularly high levels of mRNA for CRH-R2 are evident within the lateral septal nuclei, the choroid plexus, the olfactory bulb, specific amygdaloid nuclei and various hypothalamic nuclei (Chalmers *et al.*, 1995).

This heterogeneous distribution of CRH-R1 and CRH-R2 mRNA suggests a distinct functional role for each receptor in CRH-related brain systems. For example, in the anterior pituitary, the hybridisation signal for CRH-R1 mRNA is clustered around a subset of cells identified as corticotropes (Potter *et al.*, 1994, Chalmers *et al.*, 1995). In contrast, the hybridisation signal for CRH-R2 mRNA in the anterior pituitary is either undetectable (Sawchenko *et al.*, 1995) or detected only in scattered cells (Chalmers *et al.*, 1995). The selective expression of CRH-R1 mRNA within the neocortex and cerebellar cortex suggests that CRH affects in these regions are likely to be mediated by CRH-R1. On the other hand, the specific expression of CRH-R2 mRNA within the forebrain lateral septal nuclei suggests that hypothalamic CRH input to this region is likely to be mediated by CRH-R2. Such anatomical information provides a basis for functional hypotheses related to CRH-R subtypes. The CRH-R1 may be regarded as the primary neuroendocrine pituitary CRH-R and important in cortical, cerebellar and sensory roles of CRH. The anatomical distribution of CRH-R2 mRNA indicates a role for this novel receptor in neuroendocrine, autonomic and general behavioural actions of central CRH (Chalmers *et al* ., 1995).

The CRH-R2α appears to be the predominant CRH-R2 isoform which is expressed on neuronal tissue, as where CRH-R2β is localised to non-neuronal elements such as the choroid plexus and cerebral arterioles. The localisation of CRH-R2 to cerebral arterioles (Lovenberg *et al.*, 1995b) provides a putative means by which CRH may act to directly modulate cerebral blood flow. In peripheral organs, CRH-R2α and CRH-R2β mRNA are co-expressed in cardiac and skeletal muscle (Valdenaire *et al.*, 1997). CRH-R2β is also present in the lung and intestine (Perrin *et al.*, 1995). CRH-R2γ, which has been isolated only from humans, was found only in the brain (Kostich *et al.*, 1998). In human brain, CRH-R2γ mRNA expression in septum and hippocampus with weaker expression in amygdala, nuleus accumbens, midbrain and frontal cortex. CRH-R1α has been detected in the skin (Slominski *et al.*, 1995) and ovary (Nappi and Rivest, 1995). CRH-R1c has been characterised from the hippocampus and cerebellum of a normal 2-year old female. The CRH-R1c differs from CRH-R1α only in the deletion of 40-amino acids from the N-terminal domain predicted to extend extracellularly (Ross *et al.*, 1994).

1.4 Placental CRH

At present, the initiation of parturition in humans remains an enigma. It is probable that the mechanism controlling uterine transition from relaxation to the contractile stage is multifactorial. Two important factors are oxytocin, which is important for the initiation of labour (Fuchs *et al.*, 1982), and prostaglandins, which appear to be essential for the progression of labour (Karim *et al.*, 1970).

Recently, interest has been directed toward a third factor, placentally derived CRH. The placenta is an important source of circulating CRH (Perkins and Linton, 1995). CRH is secreted by placental trophoblast into the maternal circulation. In maternal plasma the levels of CRH increase exponentially as pregnancy progresses (Campbell *et al.*,1987) and reach concentrations 1000 fold greater than those found in the plasma of non-pregnant women. CRH has been implicated as a potential regulator of human parturition (McLean *et al.*, 1994).

The bioactivity of circulating CRH is also known to be influenced by the CRH binding protein (CRH-BP) (Potter *et al.*,1991), which binds to the hormone in an equimolar ratio and prevents its recognition at the CRH-R. It has been previously shown that CRH-BP is present in the maternal plasma during pregnancy (Linton *et al.*, 1993), but the relative concentrations of CRH and CRH-BP have not been reported, and the bioactivity of circulating CRH was therefore uncertain. Furthermore, it has been recently reported (McLean *et al.*, 1994) that the exponential rise in maternal plasma CRH concentrations with advancing pregnancy is associated with a concomitant fall in concentrations of the specific CRH binding protein in late pregnancy, leading to a rapid increase in circulating levels of bioavailable CRH at the time that coincides with the onset of parturition. In particular it has been suggested that there is a " CRH placental clock" which is active from the early stages of human pregnancy and determines the length of gestation and the timing of parturition and delivery (McLean *et al.*, 1995).

The biological role of this placental CRH is still uncertain; It may influence placental ACTH secretion (Petraglia *et al.*, 1987) although maternal plasma ACTH level does not rise proportionally (Rees *et al.*, 1975). CRH is a potent vasodilator of the human

foeto-placental circulation and may act as a regulator of placental vascular tone (Clifton *et al.*, 1994). CRH is able to prime and potentate the effects of oxytocin (Quartero *et al.*, 1992) and $PGF_{2\alpha}$ (Benedetto *et al.*, 1994) on the contraction force in human pregnant myometrium *in vitro*, suggesting that placental CRH may act as an important modulator of the human myometrium. Several reports indicate that CRH may stimulate prostaglandin production in placenta, decidua and fetal membranes *in vitro* (Jone and Challis, 1989); $PGF_{2\alpha}$ and PGE_2, in turn, are able to induce synthesis of CRH by human placenta (Petraglia *et al.*, 1987).

1.5 Myometrial CRH-Rs

The identification of specific CRH-Rs in human myometrium suggests that placental CRH might influence myometrial function, and hence parturition. Evidence for a direct action of CRH on human myometrium was provided by the identification of a single, high affinity population of CRH-Rs in the human myometrium which increase their affinity during the latter stages of pregnancy (Hillhouse *et al.*,1993). In contrast, in non-pregnant myometrium, these CRH-Rs exist in a low affinity state which is not functionally coupled to adenylate cyclase. As pregnancy progresses, the CRH-R adopts a high affinity state and becomes functionally coupled to both adenylate cyclase and cyclooxygenase (Grammatopoulos *et al.*, 1994). The functional significance of these events remain to be determined but it is likely that they promote myometrial relaxation since cAMP and PGE_2 have been shown to be potent myometrial relaxants (Wikland *et al.*, 1982). Hillhouse and Grammatopoulos suggested that during pregnancy CRH play a "protective" role for the myometrium, by preventing uterine contractions. In human pregnant myometrium at term, there is a modification in the coupling mechanisms between CRH-Rs and the catalytic component of adenylate cyclase, resulting in a reduction of CRH-stimulated cAMP production (Grammatopoulos *et al.*, 1996). As term approaches, these changes appear to reduce myometrial relaxation and facilitate the action of contractile agents such as oxytocin and $PGF_{2\alpha}$. Oxytocin may play a role here because it can exert a direct inhibitory influence on myometrial adenylate cyclase activity.

In pregnant (at term but not pre-term) myometrium, oxytocin can switch the CRH-R to a low affinity state and uncouple it from adenylate cyclase; an affect mediated via activation of protein kinase C, leading to phosphorylation and subsequent desensitization of the CRH-R (Grammatopoulos and Hillhouse, 1999). This oxytocin effect appears to be specific to the human myometrium and is mediated by two basic isoforms of the CRH-R which are expressed in the human myometrium.

1.6 CRH-R genomic structure and chromosomal localisation

Proteins were originally thought to be encoded by continuous segments of DNA, but actually very few genes contain uninterrupted coding regions. The vast majority of genes are interrupted by one or, typically, several noncoding regions called intervening sequences or introns. Introns are transcribed into RNA but are removed when it is processed into its mature form (mRNA); therefore, the sequences contained in the intron are not represented in the translated product. The coding sequences, called exons, actually dictate the protein sequence. As illustrated in Figure 1.8, genes not only include the coding region but also include surrounding sequences that may help in regulation of gene expression. These sequences include: "start" and "stop" signals for both transcription and translation; the promoter region, which lies 5' to the coding region and helps regulate transcription; the 5'-untranslated region (5'-UTR), which may function in regulation; and the 3'-untranslated region (3'-UTR), which contains the polyadenylation signal that is important for maturation of mRNA and also may be involved in other aspects of RNA processing, transport, degradation, and translation.

It is important to realise that for eukaryotes the active gene unit contains not only all the direct information defining the protein sequence as encoded in the mRNA and the sites to directly initiate and terminate mRNA transcription, but all of the information necessary to modulate gene transcription as well. Such information is represented in specific DNA sequences that act as sites to bind nuclear transcription factors that alter gene activity.

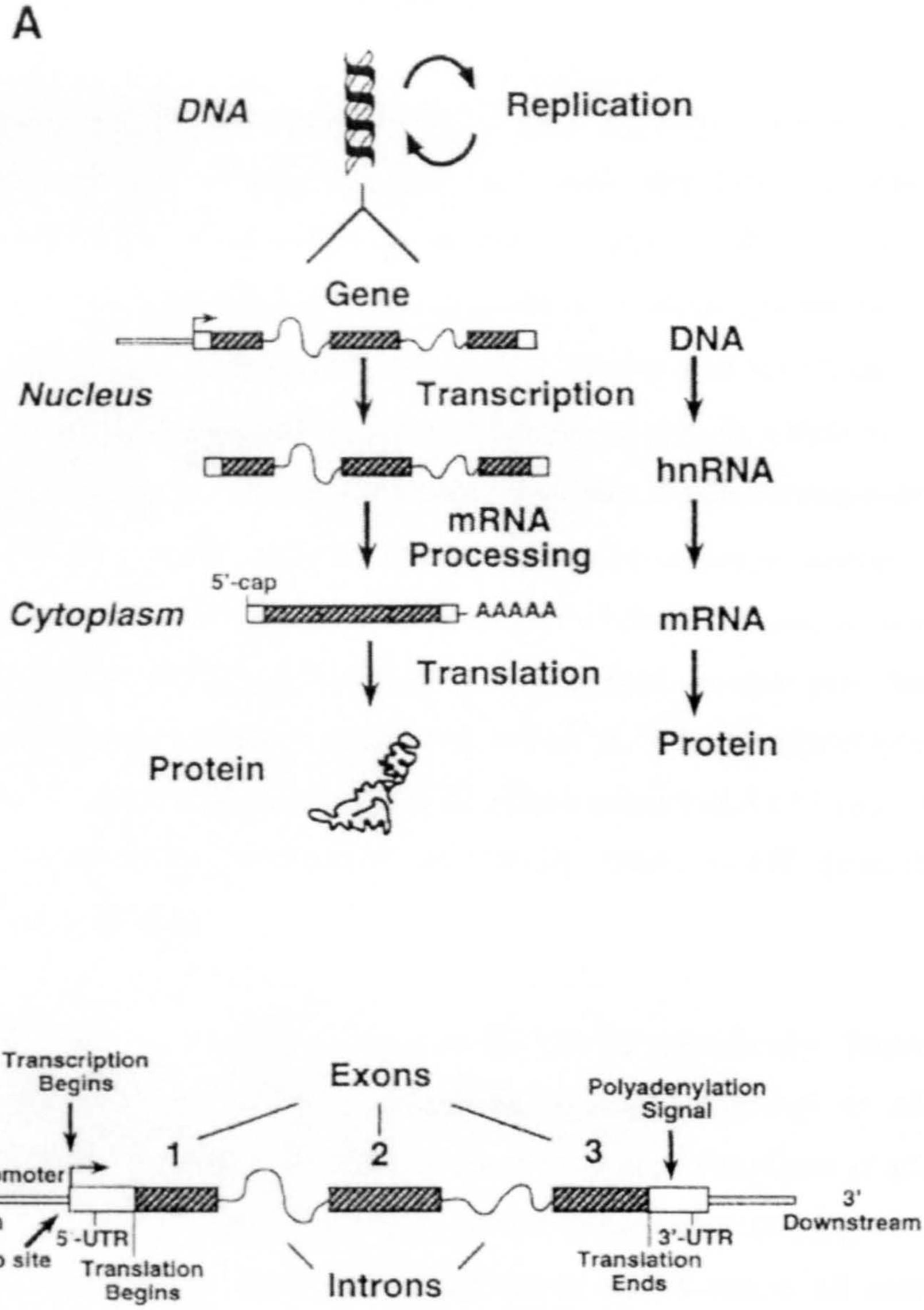


Figure 1.8. Gene structure and organisation.

(A) DNA is replicated and transcribed into RNA, which is then processed to remove introns. Translation then produces a functional protein. (B) Diagram of the basic aspects of a gene for proper transcription and translation of the DNA (Elsea, S. H and Patel, P. 1998).

1.6. 1 The genomic structure of the human CRH-R1

The gene coding for CRH-R1 consists of at least 14 exons and spans over 20 kilobases (Sakai *et al.*, 1998). The CRH-R1 gene has been mapped to the chromosome 17q12-q22 (Vamvakopoulos and Sioutopoulou, 1994). CRH-R1's four reported isoforms originate from the same gene by alternative splicing (Figure 1.1). CRH-R1α, which binds to CRH with the highest affinity and transduces the most sensitive cAMP accumulation in response to CRH, is encoded in a total of 13 exons. CRH-R1β contains an additional 29 amino acid sequence which corresponds to exon 6. Unlike the CRH-R1α, CRH-R1c lacks a 40 amino acid sequence, corresponding to exon 3. Exon-intron boundaries are the same as that of the consensus sequence. The CRH-R1 gene exhibits substantial similarity to the glucagon receptor gene family that is characterised by the presence of introns within its trasmembrane/cytoplasmic module, highly conserved cysteines in its extracellular domain and a highly conserved first intracellular loop. All members of this family induce cAMP generation and phosphoinositide hydrolysis.

Several variant receptors have been found in the GPCR superfamily. Three human calcitonin receptor (CT-R) isoforms have been reported (Albrandt *et al.*, 1995, Kuestner *et al*, 1994, Egerton *et al.*, 1995; Nakamura *et al.*, 1995, Gorn *et al.*, 1995). Differences in the three isoforms result from distinct transcription starting sites and an alternative splicing manner. Only the largest CT-R cDNA has a 16 amino acid insertion in the first putative intracellular domain of the deduced protein. This insertion is located between exon 5 and exon 7 of the hCRH-R1 gene. In growth hormone-producing pituitary adenomas, a variant growth hormone releasing factor receptor (GRF-R) mRNA has been reported (Hashimoto *et al.*, 1995). This variant has a 561 base pair insertion which contains a premature stop codon in-frame. This cassette is inserted at the position that aligns with the junction of hCRH-R1 exons 11 and 12. Rat and human type I adenylate cyclase-activating polypeptide receptor (PACAP-R) have been shown to express four major splice variants due to alternative splicing of two exons and to exhibit coupling to adenylate cyclase and phospholipase C (Spengler *et al.*, 1993, Pisegna and Wank, 1996). The insertion of the two exons in the third putative intracelluar domain is located at the junction of hCRH-R1 exons 11

and 12. The existence of these variants indicates that there may be other unknown splice variants of hCRH-R1. The 5'-flanking region of rat CRH-R1 gene was isolated and characterised (Tsai-Morris *et al.*, 1996). So far, the sequence of the promoter/5'-flank and transcription initiation site of the human CRH-R1 gene has not been reported.

1.6.2 The genomic structure of human CRH-R2

The gene coding for human CRH-R2 consists of at least 12 exons and spans approximately 30 kilobases (Liaw *et al.*, 1996). The gene has been mapped to the chromosome 7p21-p15 (Meyer *et al.*, 1997). Analysis of the human CRH-R2 gene reveals a genomic structure similar to that of mouse CRH-R1, which has 12 introns, the last 10 of which interrupt the coding region at identical positions to human CRH-R2 gene introns. The mouse CRH-R1 gene has one extra intron (intron 1) in the most 5'-coding region. Intron 2 of mouse CRH-R1 and intron 1 of human CRH-R2 also occur at different positions of the coding region. A splice variant for human CRH-R1β has 29 amino acids inserted in the first intracellular loop. The insertion occurs exactly at the corresponding position of the exon 4 and 5 junction of the human CRH-R2. There is no structural counterpart of this CRH-R1 splice variant in CRH-R2. Human CRH-2α and rat CRH-R1 genes contain 13 and 12 exons, respectively. The N-terminal extracellular domains of human and rat CRH-R1 are divided with four introns, whereas the N-terminal extracellular domain of human CRH-R2α was separated by three introns. However, in the domains toward the C-terminus, the locations of introns are quite similar among the three genes.

Molecular cloning techniques have identified three alternatively spliced forms of the CRH-R2 differing in their N-terminal domains, which are CRH-R2α, CRH-R2β and CRH-R2γ (Figure 1.10). The genomic structure of the human CRH-R2 gene was described (Liaw *et al.*, 1996), but this did not extend 5' of the CRH-R2 alternative exon. Recently Walter *et al* date reported the genomic order of the alternative exons to be β-γ-α. The CRH-R2β and CRH-R2γ exons possess good consensus splice junctions flanking an apparent 2.9-kb intron. The distance between the CRH-R2γ-

exon and the CRH-R2α-exon is not known, but it is at least 5.8 kb and may be significantly more (Kostich *et al.*, 1998).

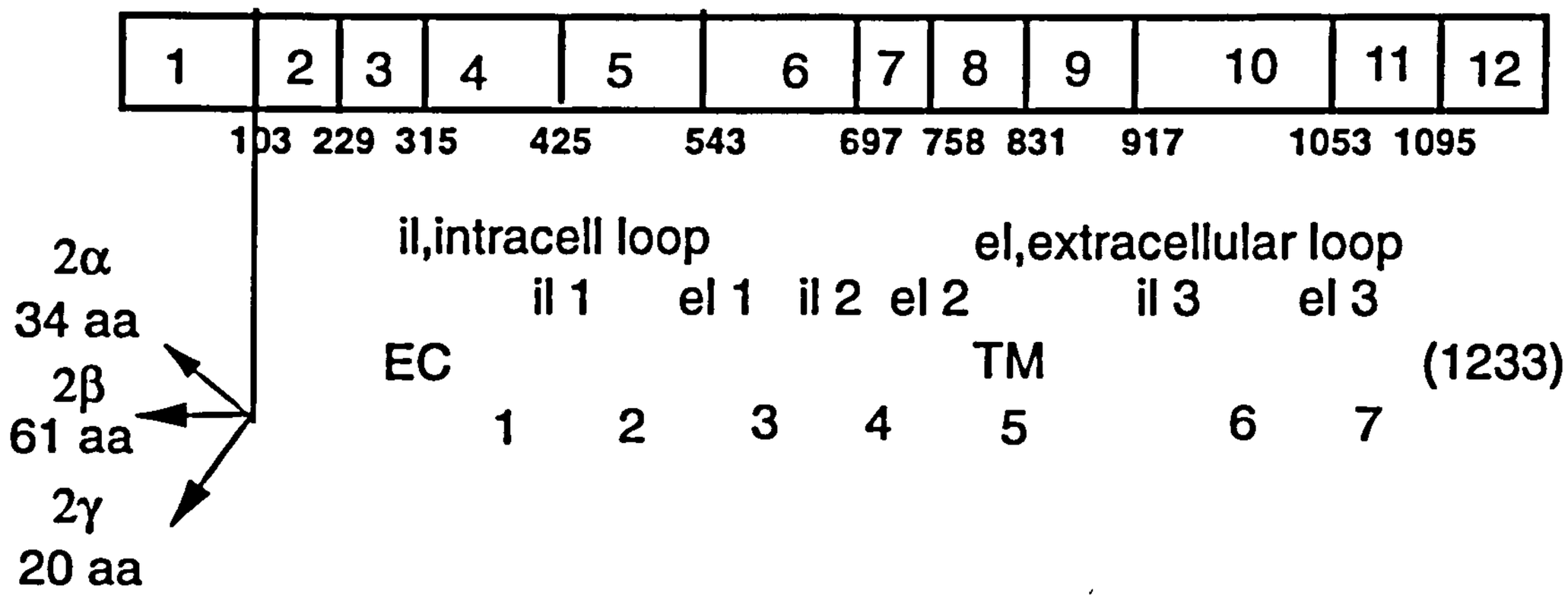


Figure 1.9. Genomic structure of human CRH-R2

The figure shows the position of CRH-R2 introns in relation to extracellular and transmembrane domains, where EC, extracellular; il, interacellular loop; el, extracellular loop; cytoplasmic domain after TMVII (1-1233 bp): initiation to terminating codons. The numbers under the boxes refer to nucleotide sequence numbers (hCRH-R2α: GenBank Accession No. U34587). The gene coding for human CRH-R2 consists of at least 12 exons and spans approximately 30 kilobases (Liaw *et al.*, 1996). extracellular (EC); intracellularloop (il); extracellular loop (el); transmembrane(TM) domains;

1.7 Eukaryotic gene regulation and promoter structure

Regulation of eukaryotic gene expression can potentially occur at many levels. These include (1) transcriptional control; (2) processing of the pre-mRNA; (3) transport to the cytoplasm; (4) stability of the mRNA; (5) selecting which mRNAs are translated; and (6) post-translational modification of the protein product. Most eukaryotic genes are regulated in part at the transcriptional level. Transcriptional control is the primary mechanism controlling gene expression. Transcription is modulated by the interaction

of regulatory molecules with short DNA sequences most of which are located upstream from affected genes. In the following section, emphasis is placed on transcriptional control mechanisms. The typical eukaryotic gene consists of up to four distinct transcriptional control elements. These are: firstly, the promoter itself; secondly, upstream promoter elements located close to it which are required for efficient transcription in any cell type; thirdly, other elements adjacent to the promoter which are interdigitated with the upstream promoter elements (UPEs) and which activate the gene in particular tissues or in response to particular stimuli; and finally, more distant enhancer elements which increase gene activity either in all tissues or in a regulated manner. Such sequences often act by binding positively acting transcription factors which then stimulate transcription. However, although most sequences act in such a positive way, some sequences do appear to act in a negative manner to inhibit transcription.

The internal structure of eukaryotic genes has been discussed in Section 1.6. In addition, there are regulatory sequences adjacent to genes that control transcription. These sequences are of two types: promoters and enhancers. These regulatory elements can be on either side of a gene, or at some distance from the gene. They are called *cis* regulators because they are found adjacent to the structural genes they regulate, as opposed to *trans* regulators (e.g., binding proteins), which are not adjacent to the genes they regulate.

1.7.1 Promoter structure

The promoter region of most genes contains several elements. The first is a sequence called the TATA box, which is located at about 25 to 30 bases upstream from the initial site of transcription (designated as -25 to -30). Some promoters have a second short conserved sequence, instead of or together with a TATA box, which surrounds the start site and is called the Initiator (Inr) element (O'Shea-Greenfield and Smale, 1992). Core promoter structures can contain both elements (composite: $TATA^{+}$ Inr^{+}), one element (distinct: containing either $TATA^{+}$ Inr^{-} or $TATA^{-}$ Inr^{+}) or no elements (null: $TATA^{-}$ Inr^{-}). Composite core promoters are found primarily in viral genes. The cellular $TATA^{+}$ Inr^{-} promoter appears to be more common than its counterpart, the

TATA$^-$ Inr$^+$ promoter, although the latter is prevalent among the hematopoetic lineage-specific genes and the homeotic genes in *Drosophila* and mammals (Carl *et al.*, 1996). The TATA$^-$ Inr$^-$ promoters might be involved in housekeeping functions, although their functional distribution is not well known. The TATA box is found in most eukaryotic genes, but it is absent in some genes, notably housekeeping genes expressed in all tissues and in some tissue specific genes (Weis and Reinberg, 1992), where it is replaced by a sequence known as the initiator element, which is located over the start site of transcription. This appears to play a critical role in determining the initiation point and acts as a minimal promoter capable of producing basal levels of transcription.

In promoters which contain a TATA box and in those which lack it, the very low activity of the promoter itself is dramatically increased by other elements located upstream of the promoter. These elements are found in a very wide variety of genes with different patterns of expression indicating that they play a role in stimulating the constitutive activity of promoters. The basal transcription of a eukaryotic gene is controlled by a number of common elements (transcription factors). Two of the best characterised transcription factors responsible for basal transcription are the CCAAT box and a GC-rich sequence known as the SP-1 box. These sequences play an essential role in efficient transcription of the gene and have been termed upstream promoter elements (UPE).

It has been reported that the 5'-end of vertebrate genes are often associated with an increase in (G+C) content, in particular with an increase in the frequency of the dinucleotide CpG (Bird, 1986). These CpG-rich islands have an average G+C content of ~60%, compared with the 40% average in bulk DNA. They take the form of stretches of DNA several hundred base pairs long. In some cases, CpG-rich islands begin just upstream of a promoter and extend downstream into the transcribed region before petering out. It is possible that the presence of unmethylated CpG-rich islands is connected with constitutive gene expression. Some times, the presence of such islands is taken as an indication that an associated sequence comprises an active gene. According to Bird (1986), there is no relationship between the presence or absence of TATA boxes and CpG-rich 5'-DNA regions.

1.7.2 Enhancers

One of the characteristic features of eukaryotic gene expression is the existence of sequence elements located at great distances from the start site of transcription which can influence the level of gene expression. These elements can be located upstream, downstream or within a transcription unit, and function in either orientation relative to the start site of transcription. They act by increasing the activity of a promoter, although they lack promoter activity themselves and are hence referred to as enhancers (Muller *et al.*, 1988). Some enhancers are active in all tissues and increase the activity of a promoter in all cell types whilst others function as tissue-specific enhancers which activate a particular promoter only in a specific cell type.

As with promoter elements, enhancers contain multiple binding sites for transcription factors which interact together (Carey, 1998). In many cases these elements are identical to those contained immediately upstream of gene promoters. Enhancers consist of sequence elements which are also present in similarly regulated promoters and may be found within the enhancer associated with other control elements or in multiple copies.

Promoter sequences are essential for transcription, while enhancers are not. Enhancers are able to stimulate levels of transcription at a distance, whereas this is not true for promoters. In addition, enhancers are responsible for tissue-specific gene expression. As a group, promoters contain modules (the TATA box) that must be located at a fixed site upstream of the transcription start site, but enhancers do not contain these modules. Although enhancers and promoters share similar sequences, the sequences in enhancers are most often contiguous. In promoters, they are spaced apart.

1.7.3 Transcription factors

The presence of suitable DNA elements (e.g. TATA/CCAAT boxes) and RNA polymerases is not sufficient to initiate transcription of most eukaryotic genes.

Proteins, that are not part of the RNA polymerase molecule, are needed for the initiation of transcription and these are called transcription factors.

The anatomy of a promoter is usually defined by a combination of gene transfer experiments to assess the effect of promoter mutants and studies of protein-DNA interactions. In the gene transfer experiments, various promoter regions are linked to a reporter gene, such as the gene encoding chloramphenicol acetyl transferase (CAT), β-galactosidase or luciferase. Sequential deletions of the promoter provide a gross delineation of the locations of regulatory elements, but this strategy may miss functional elements that act in combination with sequences that have been deleted. Ideally, point mutations are introduced into candidate regulatory elements to assess their role in the context of the native promoter. It is advantageous to simultaneously search for transcription factor binding sites, either using DNase I footprinting or gel shift assays.

A large number of transcription factors have now been identified in eukaryotic cells. Some of these bind to enhancer regions. Others bind to *cis*-acting elements in the promoter. *Trans*-acting factors (TAFs) help RNA polymerase II to initiate transcription and together with RNA polymerase they form a preinitiation complex. Many transcription factors have now been sufficiently purified to enable identification and assignment to families based on DNA-sequence binding similarities. Such transcription factors are extensively reviewed elsewhere (Latchman, 1997).

1.7.3.1 Transcription factors and inducible gene expression

All cells respond to various treatments by activating or repressing the expression of particular genes. There are a number of consensuses DNA segments which blind transcription factors (Table 1.1).

Inspection of the regulatory regions of genes which showed similar patterns of transcription, revealed the presence of short homologous DNA sequences which were absent from other genes which did not show this pattern of regulation. Thus, for example, genes whose transcription is induced in response to exposure to elevated

temperature contain a common regulatory element known as the heat shock element (HSE), which is absent from genes that do not show heat inducible transcription (Davidson *et al.*, 1983). Similarly, genes whose transcription is induced by exposure to glucocorticoid hormone contain a common glucocorticoid response element (GRE), which is absent in genes that do not show this pattern of induction (Latchman *et al.*, 1997).

Table 1.1 Sequences that confer response to a particular stimulus

Consensus sequence	Response to	Protein factor	Examples of genes containing sequences
CTNGAATNTT CTAGA	Heat	Heat-shock transcription factor	*hsp70, hsp83,* **hsp27**
T/G T/A CGTCA	Cyclic AMP	CREB/ATF	Somatostatin fibronectin, α-gonadotrophin, *c-fos, hsp70*
TGAGTCAG	Phorbol esters	AP-1	Metallothionein IIA, α_1-antitrypsin, collagenase
$CC(A/T)_6GG$	Growth factor in serum	Serum response factor	*c-fos, Xenopus* γ-actin
RGRACNNN TGTYCY	Glucocorticoid	Glucocorticoid receptors	Metallothionein II A, tryptophan oxygenase
RGGTCANNN TGACCY	Oestrogen	Oestrogen receptor	Ovalbumin, conalbumin, vitellogenin
RGGTCAT GACCY	Thyroid hormone retinoic acid	Thyroid hormone receptors	Growth hormone, myosin heavy chain
TGCGCCCGCC	Heavy metals	Mep-1	Metallothionein genes
AGTTTCNN	Interferon-α	Stat-1	Oligo A synthetase
TTTCNC/T		Stat-2	Guanylate-binding protein
TTNCNNNAA	Interferon-γ	Stat-1	Guanylate-bindig protein,Fc

Remarks: CREB/ATF represents cAMP-responsive element binding protein, AP-1 represents activator protein 1, Stat-1 represents signal transducers and activators of transcription (STAT) protein. N indicates that any base can be present at that position, R indicates a purine, i.e. A or G, Y indicates a pyrimidine, i.e. C or T.

1.7.3.2 Transcription factors and cell type-specific transcription

In eukaryotes, transcription factors also play a critical role in processes which have no parallel in prokaryotes. The higher eukaryotes contain a vast range of different cell types, each of which expresses specific genes encoding particular products necessary for the specialised function of that cell type. The role of transcription factors in controlling the cell type-specific expression of particular genes is the subject of this section.

Regulation of cell type-specific transcription is via the activity of specific transcription factors, which are synthesised or are present in an active form only in the particular cell type where the genes are active. These factors may bind directly to specific DNA sequences in their target genes as occurs for Oct-2 or NFκB or, like OCA-B, may act as co-activators which are recruited to the gene by DNA-bound factors which, like Oct-1, may be expressed in all cell types. By interacting both with each other and constitutively expressed factors, these cell type-specific factors control the specific transcription pattern of the genes which are dependent upon them.

1.7.3.3 DNA binding by transcription factors

Detailed analysis of a number of different transcription factors has indicated that they have a modular structure in which specific regions of the molecule are responsible for binding to the DNA, whilst other regions produce a stimulatory or inhibitory effect on transcription. Studies on the DNA binding regions of different transcription factors have revealed several distinct structural elements which can produce DNA binding (Harrison, 1993).

Indeed, transcription factors are frequently classified on the basis of their DNA binding domains and a selection of these binding domains is listed in Table 1.2. Well characterised DNA binding domains include: the helix-turn-helix motif found in the homeobox transcription factors (Kornberg, 1993); the two cysteine-two histidine zinc finger which is found, for example, in the Sp transcription factor family (Lania *et*

al.,1997); the multi-cysteine zinc finger which is found in the steroid-thyroid hormone receptor family (Parker, M. G 1993).

This last example is of particular interest since factors containing the basic DNA binding domain can only bind to DNA once they have formed transcription factor dimers. Hence, factors containing the basic binding domain are further sub-grouped according to the nature of the dimerization motif. Thus, some of these factors contain a helix-loop-helix motif which mediates dimerization (Littlewood and Evan, 1995). In contrast, other basic DNA binding domain-containing factors, such as the CREB factor, undergo dimerization via the so-called leucine zipper motif which contains a regular array of leucine residues (Hurst *et al.*, 1996). Thus, a wide variety of DNA binding domains (which in some cases have associated dimerzation domains) allow transcription factors to bind to their appropriate DNA sequences within target genes.

Table 1.2 Transcription factors families classified by their DNA binding domains

Domain	Factors containing domain	Comments
Homeobox	Numerous Drosophila homeotic genes,related genes in other organisms	DNA binding mediated via helix-turn-helix motif
POU	Oct-1, Oct-2, Pit-1, Unc-86	Consists of POU-specific domain and POU-homeobox
Paired box	Various Drosphila segmentation genes. PAX factors	Often found together with a homeobox in PAX factors
Cysteine-histidine motif zinc finger	TFIIIA, Kruppel, Sp1, etc.	Multiple copies of finger
Cysteine-cysteine zinc finger	Steroid-thyriod hormone receptor family	Single pairs of fingers. Related motifs in Adenovirus E1A and yeast GAL4, etc.
Basic domain	C/EBP, c-fos, c-jun, c-myc, MyoD, GCN4	Often found in association with leucine zipper or helix-loop-helix dimerization motifs
Winged HTH	Fork head, HNF 3A c-ets c-erg, Drosophila E74, PU.1	Binds purine-rich sequences
Ets domain	Ets-1, Elk-1, SAP	Contain helix-turn-helix motif

In addition to the DNA binding domain, many transcription factors contain specific regions which are necessary for the activation of transcription. As with DNA binding

domains, a number of distinct types of activation domains have been identified which are defined on the basis that they are rich in acidic amino acids, glutamine residues or proline residues, respectively. Following binding to their appropriate DNA binding site the activation domains of specific activating transcription factors can interact with the basal transcriptional complex so as to stimulate transcription. In this manner, the binding of specific transcription factors can stimulate gene transcription.

1.7.4 Transcriptional regulation of eukaryotic cells by second messenger signalling pathways

The ability of cAMP to stimulate expression of a variety of genes suggested that it might work through a discrete element that would be shared by a family of responsive genes. Activation of adenylate cyclase by various effector systems elevates intracelluar cAMP and stimulates protein phosphorylation by cAMP dependent protein kinases and transactivation of genes with the consensus palindromic sequence 5'-TGACGTCA-3' in the 5'-flanking regions (Deutsch *et al.*,1988). Several genes which are regulated by a variety of stimuli contain similar sequences in their promoter regions, although at different positions. A comparison of the CRE (cAMP response element) sequences identified to date, shows that the 5'-half of the palindrome, TGACG is the best conserved (Sassone-Corsi, 1995).

The first CRE-binding factor to be characterised was CREB (CRE-binding protein; Hoeffler *et al.*, 1988) but subsequently several additional CRE-binding factors have been identified and the corresponding genes have been cloned. Most of the CRE-binding proteins were identified by screening a variety of cDNA expression libraries with CRE and ATF sites (Foulkes *et al.*, 1991). The ATF/CREB family includes several members, of which only the CREB, CREM and ATF-1 gene products have been shown to be directly phosphorylated by the cAMP-dependent protein kinase A (Sassone-Corsi, 1995). All ATF/CRE binding proteins belong to the bZip transcription factor class. Various different factors of the CREB/ATF family are able to heterodimerize with each other but only in certain combinations. A "dimerization code" exists, which seems to be a property of the leucine zipper structure of each factor.

CRE-binding proteins may act as both activators and repressors of transcription. The activators mediate transcriptional induction upon phosphorylation by PKA (Rehfuss *et al.*, 1991). Their expression is constitutive and widely distributed in various tissues in a housekeeping fashion. Among the repressors, the Inducible cAMP Early Repressor (ICER) product deserves special mention. ICER is generated from an alternative CREM promoter and is the only inducible CRE-binding protein. ICER negatively autoregulates the alternative promoter, generating a feedback loop. ICER expression is tissue specific and developmentally regulated. The kinetics of ICER expression displays the characteristics of an early response gene.

This canonical CRE is not present in the promoters of all genes that respond to cAMP. Despite the large size of the CRE binding protein family, it is apparent that other DNA regulatory elements and transcription factors are also capable of stimulating gene transcription in response to cAMP. This is not surprising once the possibility of using phosphorylation to regulate transcription is recognised. In fact, because so many transcription factors undergo phosphorylation before or during the transcription process, it may ultimately become difficult to classify regulatory elements based solely on their ability to be stimulated by cAMP.

One of the first "non-consensus" cAMP regulatory elements was identified in the human metallothionein IIA gene. The metallothionein IIA gene binds a transcription factor referred to as activator protein 2 (AP-2). An AP-2 acts as a basal level enhancer (Mitchell *et al.*,1987), but it also stimulates transcription in response to both phorbol esters and cAMP (Imagawa *et al.*,1987). AP-2 site has been identified in a number of genes, allowing derivation of the relatively variable GC-rich consensus 5'-$GCCN_3GGC$-3' (Williams and Tjian, 1991) (Table 1.3).

Table 1.3 Selected cAMP responsive DNA sequences

Gene promoter	Nucleotide sequence	References
Sequences related to canonical CREs		
Somatostatin CRE	5'TGACGTCA3'	Montminy *et al.*,1986
GPHα	5'TGACGTCA3'	Silver *et al.*,1987
VIP	5'TGACGTCT3'	Tsukada *et al.*,1987
PEPCK	5'TGACGTAA3'	Short *et al.*,1986
c-fos	5'TGACGTTT3'	Fisch *et al.*, 1989
Adenovirus ATF	5'TGACGTAG3'	Lin *et al.*, 1988
Proenkephalin	5'TGACGCAG3'	Comb *et al.*, 1986
AP-1	5'TGA-GTCA3'	Angel *et al.*, 1987
Consensus	5'TGACGTCA3'	
Other sequences involved in cAMP-regulated expression		
AP-2	5'GCCN$_3$GGC3'	Imagawa *et al.*,1987
Jun B	5'AGTGCACT3'	de Groot *et al.*,1991
Ovalbumin ERE	5'TGGGTCA3'	Doucas *et al.*, 1991
CYP11A1	5'TCACTGATGACCTTGAGCCCTGG3'	Inoue *et al* .,1991
CYP 17A	5'TTGATGGACA GTGAGCAAG3'	Lund *et al.*, 1990
CYP21B	GC-rich Sp1-like element	Kagawa *et al.*, 1990
hCGβ	GC-rich region	Albanese *et al.*, 1991
RII β	GC-rich region	Kurten *et al.*, 1992
pST-cathCP2	GC-rich region	Hjorth *et al.*, 1990

Activator protein (AP-1) was initially described as a nuclear factor required for transcription by phorbol ester tumour promoters such as TPA, an activator of the protein kinase C pathway (Lee *et al.*, 1987). The TPA response element (TRE or AP-1 site) was identified initially in the promoter region of the human metallothionein IIA and collagenase genes (Angel *et al.*, 1987). The TRE sequence has been identified in a large number of genes, that allows derivation of the consensus sequence of TGA (C or G) TCA, which differs from the palindromic CRE (TGACGTCA) by a single nucleotide. TRE or AP-1 binding proteins are the Jun and Fos families of the bZIP superfamily of transcription factors. These proteins interact through their carboxy-terminal leucine zippers, and *c-jun* can form both homodimers with itself and heterodimers with *c-fos*, while *c-fos* can heterodimerize with *c-jun* but cannot form homodimers. TPA and the activation of PKC induce *c-jun*. This is due to positive autoregulation mediated by binding of *c-jun* homodimers or *c-jun:c-fos* heterodimers to an AP-1 site within the *c-jun* promoter (Radler-Pohl *et al.*,1993, Karin M, 1990).

Although the TRE (AP-1) site was first identified based on its response to phorbol esters, it also imparts sensitivity to cAMP in JEG-3 and HeoG2 cells (Deusch *et al.*,1988). A TRE-like element in the human proenkephalin gene conveys responsiveness to both cAMP and phorbol esters (Comb *et al.,* 1988), although the response to phorbol esters alone is small and is greatly potentiated by phosphodiesterase inhibitors. The TRE mediated transcriptional responses through both the protein kinase A and protein kinase C pathways. This dual activity may reflect heterodimeric interactions between transcription factors or, alternatively, phosphorylation of TRE-binding factors by both pathways.The CRE and TRE sites have been described as "general response elements" at the distal end of a complex signal transduction pathway (Figure 1.11).

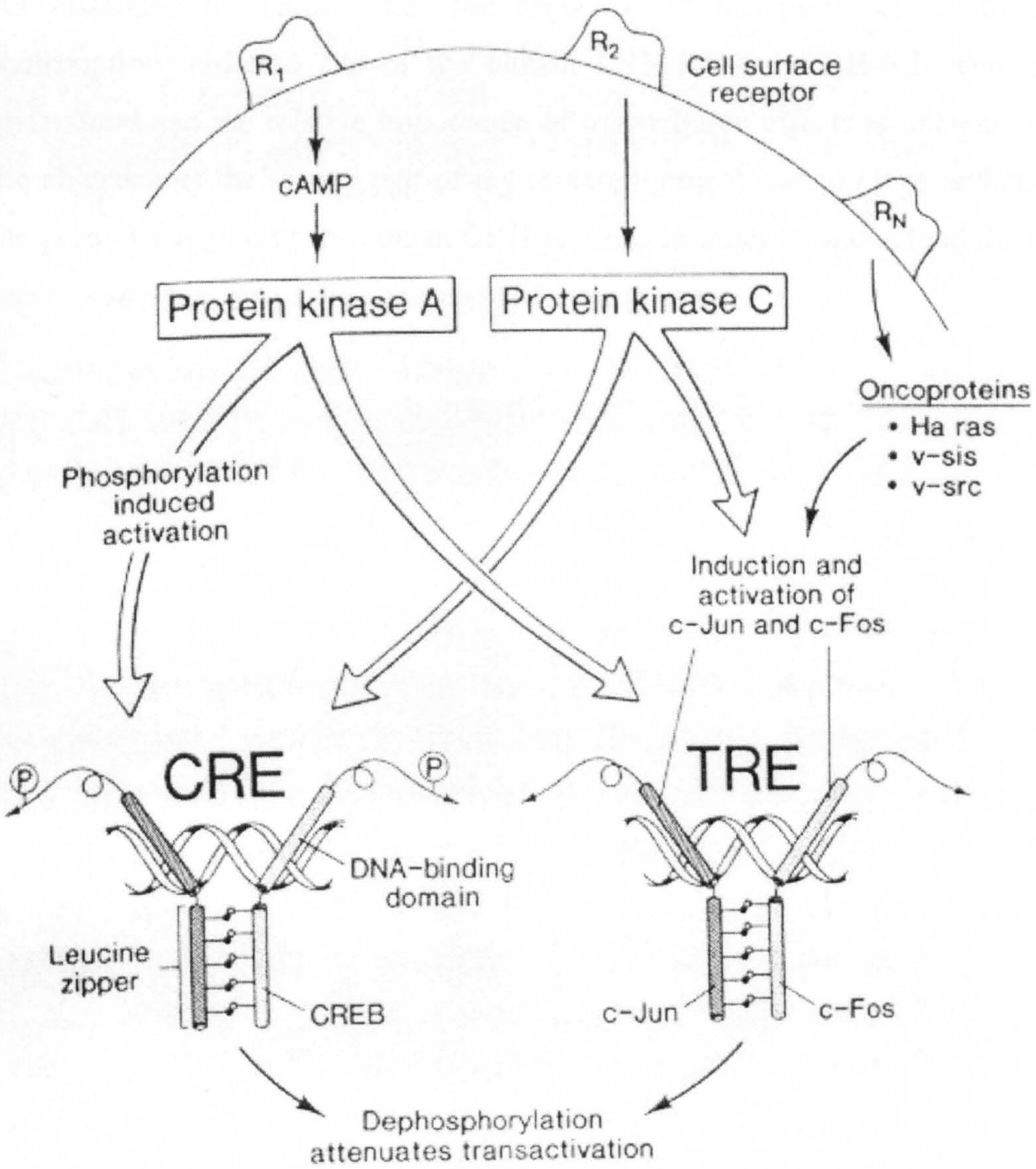


Figure 1.11. Signal transduction pathways that modulate gene transcription.

Several examples of signal transduction "cross-talk" suggest that second messenger response elements can be activated by either the Protein kinase A or Protein kinase C pathway (Hunter T and Karin M 1992).

1.8 Objective of the research project

The objective of the first part of my research project was to detect the presence of CRH-R subtypes in human myometrial biopsies (pregnant and non-pregnant) and myometrial cell cultures.

As discussed in Section 1.3, the sequence of the promoter/ 5'-flanking and transcription initiation site of the human CRH-R1 and CRH-R2 gene is not yet understood and the relative importance of transcription effects is unclear. Therefore, the objective of the second part of my research project was to clone and characterise the promoter region of the human CRH-R genes in order to understand the molecular mechanisms that direct the expression of these genes.

Chapter 2 Materials and Methods

The majority of materials used in the experiments related to the project are described in this chapter, which may include the names of the materials, suppliers, and general applications. The experimental methods or techniques adopted in the project are also briefly described in this chapter. The chapter is organised in two sections, namely, Materials and Methods.

2.1 Materials

2.1.1 General reagents

Restriction enzymes were obtained from Gibco BRL and New England Biolabs, all other enzymes were obtained as stated in the text.

All radioisotopes were supplied by Amersham at the following specific activities:

[α-^{32}P] CTP 800 Ci/mmol;

[α-^{32}P] dCTP 3,000 Ci/mmol;

[γ-^{32}P]ATP3,000Ci/mmol;

[α-^{35}S]dATP 1,000Ci/mmol.

X-ray films were from Fuji and Kodak.

Oligonucleotides were synthesised by the University of Warwick Oligonucleotide Synthesis Core Facility (Applied Biosystems DNA synthesiser, model 380B) or were ordered from Gibco BRL.

Material for bacterial growth media were supplied by Difco laboratories (Michigan, USA) and Oxoid Ltd. (England). Antibiotics were supplied by Sigma Chemical Co., St. Loius, USA. Petri dishes were supplied by Falcon.

Agarose were supplied by Gibco BRL. Acrylamide was supplied by Fisons and bisacrylamide by Kodak. The other chemicals were from BDH, Sigma, or Fisons.

Molecular marker: 1kb DNA ladder was from Gibco BRL.

2.1.2 Stock solutions

TE: 10mM Tris-Cl, pH 7.4 EDTA, pH 8.0.

10 x TBE: 0.9 M Tris-borate, 20 mM EDTA, pH8.0.

20 x SSC: 3M NaCl, 0.3M sodium citrate, pH7.0.

10 x MOPS: 0.1M MOPS, pH7.0, 40mM sodium acetate, 5mM EDTA, pH 8.0.

Phosphate buffered saline (PBS): 0.8% (w/v) NaCl, 0.02% (w/v) KCl, 0.144% (w/v) Na_2HPO_4, 0.024% (w/v) KH_2PO_4, pH 7.4.

2.1.3 Bacteriological media

LB broth: 0.5% (w/v) bacto-yeast extract, 1.0% (w/v) bacto-tryptone, 1.0% (w/v) NaCl, pH 7.0.

2 x YT broth: 1.6% (w/v) bacto-tryptone, 1.0% (w/v) bacto-yeast extract, 0.5% NaCl, pH7.0

Agar was added to 1.5% (w/v) for pouring plates.

Antibiotics were used in plates and broth at the following final concentrations: ampicillin, 50-100µg/ml; tetracyclin, 15µg/ml.

2.1.4 Cell lines

The HEK-293 cell line was obtained from the ECACC (European Collection of Animal Cell Cultures) in frozen ampoules. Myometrial cell line is derived from a primary culture of uterine myocytes.

2.1.5 Tissue collection

Pregnant myometrial biopsies were obtained from women undergoing elective caesarean section at term. The biopsy site was standardised to the upper margin of the lower segment of the uterus in the midline, which provides the closest approximation to the upper segment of the uterus.

Non-pregnant myometrial tissues were obtained from premenopausal controls undergoing hysterectomy for non-malignant conditions. The biopsies were immediately snap-frozen in liquid nitrogen and subsequently stored at - 70^0C until use. Ethical approval was obtained from the local ethical committee and informed consent to the study was obtained from all patients.

2.1.6. Plasmid vector

pBluescript II SK (+/-)(Stratagene) and pGEM-T (Promega) were used as a general purpose cloning PCR products vectors .

pGEM-5Zf(+/-) and pGEM-7Zf(+/-) were used as general purpose subcloning vector (Promega).

The pCAT3-Basic reporter gene plasmid (Promega) was used as a expression system.

2.1.7 Genotypes of *E. coli* strains

For general cloning in this project, four strains of *Escherichia coli* have been employed which had genotypes as shown below:

XL1-Blue: *recA*1, *endA*1, *gyrA*96, *thi*-1, *hsdR*17, *supE*44, *relA*1, *lac*, [F'-*proAB*, *lacI*^{q} ZΔM 15, Tn10 (tet^r)].

XL2-Blue MRF': *recA*1, *endA*1, *gyrA*96, *thi*-1, *hsdR*17, *supE*44, *relA*1, *lac*, [F' proAB *lacI*^{q}ZΔM15Tn10 (tet^r) Amy Cam]^a.

DH5α: *supE*44, *hsdR*17, *recA*1, *endA*1, *gryA*96, *thi*-1, *relA*1.

JM109: *recA*1, *supE*44, *endA*1, *hsdR*17, *gyrA*96, *relA*1, *thiΔ(lac-proAB)*.

2.1.8 PCR primers

pBluescript T3 primer: 5'-ATTAACCCTCACTAAAG-3',

pBluescript T7 primer: 5'-AATACGACTCACTATAG-3',

pBluescript Reverse primer: 5'-AACAGCTATGACCATG-3',

pBluescript M13-20 primer: 5'-GTAAAACGACGGCCAGT-3',

The PCR primers for cloning CRH-R

CRH-R1α and CRH-R1c:

1st 1S: 5'-ACCCAAGCTTAGCCGAGCGAGCCGAGGATG-3'

1A: 5'-GTCGACAAGC$(T)_{18}$-3'

2nd 2S:5'-GGGAATTCGAGGATGGGAGGGCACCCGC-3'

2A:5'-GGGAAGCTTCAGACTGCTGTGGACTGCTT-3'

CRF6: 5'-AAACAATGGCTACCGGGAGTG-3'

CRF3: 5'-TGACCAACTTCTTCTGGATGTT-3'

CRF5: 5'-CACCTACATGCTGTTCTTCGTCA-3'

CRH-R1β

1st 2S: 5'-GGGAATTCGAGGATGGGAGGGCACCCGC-3'

1A: 5'-GTCGACAAGC$(T)_{18}$-3'

2nd 5'S: 5'-CAGCTCCGTCTCGTCAAGGC-3'

5'A: 5'-GCACCGGATGCTCCTTCGAAC-3'

2nd 3'S: 5'-ATGGAGTGGTGCCCCATTTC-3'

2A: 5'-GGGAAGCTTCAGACTGCTGTGGACTGCTT-3'

CRH-R2α

1st S1: 5'-GGGAATTCCATGGACGCGGCACTGCTCC-3'

A1: 5'-GGGAATTCCCGGGCCAAGAGGCATGGTTTATTTC-3'

2nd S2: 5'-GGGAATTCTCCACAGCCTGCTGGAGGCC-3'

A2: 5'-GGGAATTCCCGGGCAGGTGGGCGACCGAGGG-3'

CRH-R2β

1st S1: 5'-TCCAGTCCCTAACCCCAGCC-3'

A1: 5'-GGGAATTCCCGGGCCAAGAGGCATGGTTTATTTC-3'

2nd S2: 5'-CTGGCATGAGGGGTCCCTCAG-3'

A2: 5'-GGGAATTCCCGGGCAGGTGGGCGACCGAGGG-3'

CRH-R2γ

1st S1: 5'-CTGTGCTCAAGCAATCTGCCTAC-3'

A1: 5'-GGGAATTCCCGGGCCAAGAGGCATGGTTTATTTC-3'

2nd S2: 5'-CTTGGCTTCCCCAAGTGCTGAG-3'

A2: 5'-GGGAATTCCCGGGCAGGTGGGCGACCGAGGG-3'

5'RACE primers used for cloning CRH-R2

CRH-R2β (1) 1st: 1A 5'-GAGAGGTTGGTGAGGGTCATGATGGTG-3'

2nd: 2A 5'-GAGGAGGTGTGGGACGTAGAGGAG-3'

(2) 1st: 1A 5'- GAGGAGGTGTGGGACGTAGAGGAG-3'

2nd: 2A 5'-GGTATTCTGGACCCACTTCTTCCAC-3'

CRH-R2α 1st: 1A 5'-CAGCCGTCCAAGAGCAGCTCTTCAG-3

2nd: 2A 5'-CTTCAGCCAGCGCCAGGCTGCAGTTG-3'

CRH-R2γ 1st: 1A 5'-GCAGAAGAGCTGAGGAAAGCCCAGTC-3'

2nd: 2A 5'-AAAGCCCAGGTCCCTGTCTTCAGGC-3'

Adaptor primer AP1: 5'-CCATCCTAATACGACTCACTATAGGC-3'

Nested Adaptor primer AP2: 5'-ACTCACTATAGGGCTCGAGCGGC-3'

5'RACE used to identify the 5'-end cDNA of CRH-R1

CRFII: 5'-GGGAAGCTTCAGACTGCTGTGGACTGCTT-3'

CRF4: 5'-GGAAGGAGTTGAAGTAGATG-3'

CRF7: 5'-ATGAACATCCATTTGCGCAGC-3'

CRF8: 5'-GTCCACGGATGCGTTGCACTGC-3'

PE1: 5'-CAGAAGGAGAAGGGCCTTGACG-3'

PE3: 5'-CTTGACGAGACGGAGCTGCGG-3'

Abridged anchor primer:

5'-GGCCACGCGTCGACTAGTACGGGI IGGGI IGGGI IG-3'

Abridged universal amplification primer (AUAP):

5'-GGCCACGCGTCGACTAGTAC-3'

Construction of plasmid for promoter analysis

P1: 5'-CCGGTACCGACCTGTGAGATTTCACAGGAG-3'

P2: 5'-CCGGTACCGAGGCTCTGGAAGAACAGACC-3'

P3: 5'-CCGGTACCACCTACCCACAGTACAGTATC-3'

P4: 5'-CGGCTCGAGAGGTGTGGGACGTAGAGGAGG-3'

PCR primers used for rat CRH-R1 promoter region

RCRF: 5'-CTAGGTGCTTCTCAACCAGGCC-3'

RCRF1: 5'-CACCTGGCCGTGCTGAACCTC-3'

Genomic walking

1) Universal-17: 5'-GAGGACTCGAGCTCAAGC$(T)_{17}$-3'

CRH-R2β (2) 1st: GAGGAGGTGTGGGACGTAFAGGAG-3'

2) Universal: 5'- GAGGACTCGAGCTCAAGC-3'

CRH-R2β (2) 2nd: 2A 5'-GGTATTCTGGACCCACTTCTTCCAC-3'

PCR primers for analysis the clone containing the exon I of CRH-R1α gene

PE5: 5'-GGAAACGGCGGCCAGACTTCC-3'

PE3: 5'-CTTGACGAGACGGAGCTGCGG-3'

PCR primers for amplify the region between the CRH-R2β and 2γ exon

B4: 5'-CATGACCCTCACCAACCTCTC-3'

G3: 5'-GGTAGGCAGATTGCTTGAGAC-3'

2.1.9 Genomic DNA library

An amplified human placental genomic cosmid DNA library was obtained from Clontech, USA. The library was constructed from 30-50 kilobase fragments of the partial *Bam*H I digest of placental DNA inserted into the *Bam*H I site of the pWE 15 cloning vector.

2.1.10 Human CRH-R1 and CRH-R2 riboprobe

Human CRH-R1 and CRH-R2 riboprobe were obtained from Dr. Lovenberg, Neurocrine Biosciences Inc., San Diego.

2.2 Methods

2.2.1 Routine manipulation of DNA

2.2.1.1 Extraction procedures

Samples containing nucleic acids were routinely subjected to phenol, phenol:chloroform: iso-amyl alcohol (25:24:1) and chloroform. In order to remove protein contaminants. In extraction procedures the sample were added to equal volumes of Tris pH 8.0 buffer phenol or phenol:chloroform:iso-amyl alcohol or chloroform only and vortexed briefly to mix thoroughly. The mixtures were subjected to centrifugation (20,000g) in a 1.5 ml microcentrifuge for 5 minutes to separate into two phases. The top aqueous layer containing the DNA was recovered (such that the interface that usually contains protein was not disturbed) and transferred into a clean microcentrifuge tube. After these procedures, the DNA was precipitated as described below.

2.2.1.2 Precipitation procedures

The salt concentration of an aqueous solution containing DNA was adjusted to 0.3 M with sodium acetate (pH 5.2) or ammonium acetate (pH 7.2), and mixed by vortexing briefly. Two and half volumes of 100% cold ethanol, or an equal volume of isopropanol, were added and then mixed thoroughly either by vortexing or by inverting the tube several times. The process of precipitation was allowed to occur either at -20°C for two hours or at -70°C for 30 minutes. The DNA was then pelleted by centrifugation (20,000 g) in a microcentrifuge for 15 minutes. The ethanol was removed, the pellets washed with 70% ethanol to remove salt, then dried briefly under vacuum or air dried. After all above procedures, the DNA pellets were resuspended in an appropriate volume of TE buffer (10mM Tris-HCl pH8.0, 1mM EDTA) or sterile distilled water.

2.2.2 Electrophoresis of nucleic acids

2.2.2.1 Non-denaturing agarose gels

Sample has 0.2 volumes of Type III loading buffer (50% v/v glycerol, 0.25% v/v bromophenol blue, 0.25%w/v xylene cyanol FF) added to them run on gels of 0.6% to 2.0% w/v agarose in 1x TBE with 0.5µg /ml ethidium bromide (from a stock solution of 10mg /ml). Gels were run in 1xTBE buffer containing 0.5µg/ ml ethidium bromide and photographed on an ultraviolet light box.

2.2.2.2 Non-denaturing polyacrylaimd gels

5% non-denaturing polyacrylamide gels were poured in gel plates 160mm in length and consisted of 5% polyacrylamide (acrylamide: *N,N'*-methylenebisacrylamide 29:1 (w/v)), 1x TBE, 0.07% (w/v) ammonium persulphate, and 0.035% (v/v) TEMED (Sigma). The gel was allowed to polymerise for 30 minutes and then the wells were flushed with 1 x TBE running buffer. DNA samples were mixed with an equal volume of 6 x gel loading buffer (50% v/v glycerol, 0.25% w/v bromophenol blue, 0.25% w/v xylene cyanol FF) and electrophoresed at 150 volt in 1 x TBE until the bromophenol blue had migrated to the bottom of the gel. Finally, the gel was dried under vacuum at 80°C and exposed to X-ray film.

2.2.2.3 Denaturing polyacrylamide gels

6% polyacrylamide (acrylamide: *N,N'*-methylenebisacrylamide 19:1 (w:v)) gels containing 8M urea, 1 x TBE, 0.05% ammonium persulphate, 0.05% TEMED (Sigma) were poured between gel plates and allowed to polymerise for 4 hours at room temperature. For electrophoresis of products from RNase protection assays, gels of 16 cm in length were used, and for separation of sequencing products gels of 40 cm in length were employed. Gels were fixed to the electrophoresis tank and pre-run with 1 x TBE running at 40 watts (Sequencing gels at 70 watts) for 30 minutes after which sample wells were flushed with running buffer.

Nucleic acid samples in denaturing gel loading buffer: 90% (v/v) formamide, 10mM EDTA, pH8.0, 0.01% bromophenol blue, 0.01% xylene cyanol, were heated at 85-

100°C for 3-5 minutes, loaded onto the gel and electrophoresed at 40 watts, for sequencing gels at 70 watts.

Sequencing gels were fixed in 10% (v/v) acetic acid for 30 minutes, transferred onto filter paper (3MM Whatman) and dried at 80°C on a vacuum drier and then exposed to X-ray film at room temperature. Gels with ^{32}P samples were transferred onto filter paper and dried at 80°C on a vacuum drier and exposed to X-ray film with an intensifying screen at - 70°C.

2.2.3 Extraction of total RNA

RNA was extracted from isolated human myometrial tissues using the RNeasy Total RNA Kit (Qiagen GMBH) according to the manufacturer's instructions.

a) The tissue was weighed, 100mg and homogenised in 600 µl lysis buffer (RAT). The flow-through fraction from QLAshredder was transferred to a 1.5ml microcentrifuge tube, and 225µl ethanol added to the flow-through fraction, and mixed by pipetting. The sample was applied to RNeasy spin column. It was centrifuged for 15 seconds at 10,000 rpm. The column was wash with 700µl washed buffer RW1, centrifuged for 1 minute, and washed again with 500µl wash buffer RPE. After centrifuged for 2 minutes to dried the spin column membrane, the RNA was eluted with 50µl water or TB buffer.

b) DNase digestion reaction. The total RNA was in combined with 5 X T3/T7 RNA buffer (Gibco-BRL), 20u RNase inhibitor, 0.1M DTT and 2 united RQ1 RNase-free DNase (Promega) in a volume of 100 µl. The sample was incubated for 37°C at 20 minutes.

c) RNeasy Total RNA Kit was used to purify RNA from DNase digestion reaction. The total RNA purity and recovery of each sample was determined using an UV spectrophotometer (260 and 280 nM). The integrity of the purified RNA was assessed by gel electrophoresis on denaturing agarose gels.

2.2.4 Isolation of genomic DNA

Tissue samples were frozen in liquid N_2 and ground to a fine powder with a pestle and mortar. The powdered tissue was transferred to a 50ml polypropylene tube (Falcon)

and 10 volumes of Extraction buffer: 10mM Tris, pH8.0, 0.1M EDTA, pH8.0, 20µg/ml pancreatic RNase (Sigma), 0.5% SDS added. The transferred tissue was incubated at 37°C for 1 hour with gentle agitation after which proteinase K was added to a final concentration of 100µg/ml and incubated at 50°C for overnight. The solution was extracted twice with an equal volume of phenol (pH8.0) and twice with an equal volume of phenol:chloroform(1:1 (v/v) pH 8.0), the aqueous phase was removed each time using a 0.5ml pipette tip with the end cut. The DNA pellet was then precipitated, washed and resuspended using the procedure as described in Section 2.2.1.2.

2.2.5 Cell culture

2.2.5.1 Growth and subculture of tissue culture cells

Adherent cells were grown in complete DMEM in disposable 75 cm^2 or 175 cm^2 tissue culture flasks (Nunclon) at 37°C in a humidified atmosphere of 5% CO_2. Confluent monolayers were wash once with PBS and incubated with prewarmed at 37°C 1 X trypsin-EDTA (5 mg/ml trypsin, 5 mM EDTA, 145 mM NaCl) in PBS at room temperature the cells started to round up and detach from the flask (1 to 3 minutes). Detachment was monitored under an inverted microscope. Then the trypsin solution was aspirated, the cells were resuspended in fresh complete medium and plate at the required density in new flasks or cell culture dishes.

2.2.5.2 Storage and recovery of cells

One 175 cm^2 flask of nearly confluent cells (approximately 10^7 cells) was trypsinized as described above. Cells were collected by centrifugation at 400 g for 5 minutes and the cell pellet resuspended in 3 ml of medium containing 50% FCS and 10% DMSO. Aliquots of 1 ml were transferred to BIO-Freeze vials (Costar). The vials were allowed to freeze slowly to - 70°C over a period of 24-48 hours and then transferred to liquid nitrogen for long-term storage.

For recovery of cells from the liquid nitrogen, the cell vials were thawed quickly in at 37°C-water bath and their contents were mixed with 10 ml of performed (37°C) complete medium. The cells were pelleted by centrifugation at 400 g for 5 minutes to

remove the toxic cryopreservative DMSO, then resuspended in 25 ml complete growth medium, transferred to a 75 cm^2 tissue culture flask and incubated under the conditions described above. Viable cells usually started to attach to the bottom of the flasks several hours after recovery from the liquid nitrogen.

2.2.6 Preparation of myometrial cell cultures

Pieces of myometrium were transferred into DMEM containing collagenase (300 U/ml), deoxyribonuclease (30 U/ml), penicillin (200 U/ml) and streptomycin (200 mg/ml), and then incubated at 37°C for 30 minutes. After filtration and centrifugation, cells were suspended in DMEM containing 10% foetal calf serum (FCS), penicillin (100 U/ml), streptomycin (100 mg/ml) and fungizone (2.5 µg/ml). The cells were kept at 37°C in a humidified atmosphere of 95% air and 5% CO_2 until confluence (2-4 weeks). The purity of myometrial muscle cells was assessed by immunocytochemical staining. Mouse antihuman smooth muscle actin-specific monoclonal antibody and peroxidase-conjugated rabbit antimouse antibody were used. Human fibroblast cells and omission of the primary antibody were used as negative controls, whereas frozen myometrial tissue was used as a positive control. To minimise fibroblast contamination we re-purified the myocyte preparation 48 hour before the experiments using 0.5% EDTA-trypsin.

2.2.7 Subcloning techniques

2.2.7.1 Restriction enzyme digests

Restriction enzyme digests were carried out according to the manufacturer's instructions. The products of a digest were analysed using agarose gel electrophoresis with the presence of DNA markers (1 kb ladder, Gibco BRL). Complete digests were checked by ensuring that uncut plasmid was not present. In the case of double digests while the optimal buffer conditions of the enzymes were different, the reaction was carried out in a buffer that gave at least 80% enzyme activity. If the reaction conditions of the two enzymes were incompatible, one digest was carried out first. Afterward, a phenol:chloroform extraction followed by an ethanol precipitation step

were carried out. The pellet was resuspended in the appropriate reaction conditions for the second enzyme and the second digest carried out as usual. When temperature requirements of the enzymes in a double digest were different, the reaction was first carried out at one temperature then transferred to the second temperature and incubated for a similar length of time.

2.2.7.2 Preparation of plasmid DNA for ligations

Vectors that had been digested with two enzymes generating incompatible ends were desalted using QIAquick spin columns (Qiagen, see Section 2.2.10.2). Vectors cut with a single restriction enzyme or blunt-ended were heat inactivated at 65°C for 10 minutes and then cooled to room temperature. For 5'-protruding ends, 1 unit calf intestinal phosphatase (CIAP) per 100 pmole DNA was added, then the enzyme was left for 60 minutes at 37°C. For 5'-recessed or blunt ends, 1 unit CIAP per 10 pmole DNA was used incubated for 15 minutes at 37°C, then for 15 minutes at 56°C. Added another aliquot of CIAP and repeat incubations at both temperatures. Finally the enzyme was inactivated with 2.0µl of 0.5M EDTA and desalted using QIAquick spin columns (Qiagen, see Section 2.2.10.2).

2.2.7.3. Making T-vector for the cloning PCR products

Certain thermostable polymerases, including *Taq, Tfl* and *Tth* DNA polymerase, add a single nucleotide, generally adenine, to the 3'-ends of amplified DNA fragments. T-vectors are convenient for the cloning of PCR products generated by such thermostable polymerases. These linearized vectors contain single 3' terminal thymidines (T's) at each end, which are complement to the overhang added by the polymerase. Thus, PCR products can be directly subcloned into these vectors without further manipulation. T-vector was made using the method as described by Mezei *et al.* (1994).

2.2.7.4 Preparation of insert DNA for ligation

Insert DNA was generally a restriction fragment isolated from an agarose gel using

QIAquick spin columns. For some purposes it is necessary to produce target DNA by PCR. This produced DNA terminus is lack of the 5' phosphate. Therefore, either the primers or the amplified DNA was phosphorylated by the addition of 1mM ATP, 1 x forward reaction buffer (70mM Tri-Cl, pH 7.6, 10mM $MgCl_2$, 100mMKCl, 1mM 2-mercaptoethanol), 20U of T4 polynucleotide kinase (Gibco-BRL) and incubated at 37°C for 30-60 minutes. The sample was heat inactivated at 90°C for 2 minutes and desalted using QIAquick spin columns (see Section 2.2.11.1).

DNA pre-digested with enzyme(s) that leave 5' overhang(s) was blunt ended by filling in the 5' overhang(s) with DNA polymerase I (Klenow fragment) in the presence of 2 mM dNTP mixture and the appropriate volume of 10 X concentrated 5' overhang buffer (0.5 M Tris-HCl pH 7.2, 0.1 M $MgSO_4$, 1 mM DTT, 500µg/ml bovine serum albumin). The reaction was allowed to proceed for 30 minutes at 37°C. The sample was heat inactivated at 70°C for 5 minutes and desalted using QIAquick spin columns (see Section 2.2.11.1).

2.2.7.5 Ligations

Bacteriophage T4 DNA ligase (Gibco-BRL) was adopted in ligations involving both blunt-ended and protruding-ended which were carried out using the method as described by Sambrook *et al.* (1989). The ligation mixtures included 100 ng vector DNA, equimolar amount of insert DNA (in some cases 1:2 or 1:3 ratio of vector:insert), appropriate amount of 5 x DNA ligase buffer (supplied with the ligase) and 1 Weiss units T4 DNA ligase in a final volume of 10-20 µl. They were incubated at 16°C for 12-16 hours. Following ligation, the samples were stored at -20°C, if they not are used immediately.

2.2.8 Preparation and transformation of *E.coil* using calcium chloride

Competent bacteria were prepared by the method described in Sambrook *et al.*, 1989 with slight modifications. A 10ml culture of LB broth and appropriate antibiotic was inoculated with a single colony of the host bacteria and incubated at 37°C overnight with shaking. They were diluted 50 x (1 ml bacterial culture in 50 ml LB broth) and incubated until the OD_{600} was around 0.3. The culture was left on ice for 30 minutes

and then centrifuged at 4°C for 10 minutes at 3,000 X g to pellet the cells. They were then resuspended in half the original volume (25ml) which was freshly prepared sterile 0.1 M $MgCl_2$ (ice cold), immediately centrifuged as before. The cell pellets were resuspended in 1/20th of the original volume (2.5ml) of ice-cold 0.1 M $CaCl_2$ (sterile, freshly prepared from solid), 15% (v/v) glycerol. The cells were then aliquoted, those to be used in transformation reactions were immediately stored for 60 minutes on ice and the rest stored at -70°C.

Freshly prepared competent cells (0.2 ml) were mixed gently with 0.1 ml DNA (1-10ng) in Tris-HCL pH 7.4 or with half of original volume of ligation mixture (5-10µl). Tubes were incubated on ice for 30 minutes with occasional mixing, and transferred at 42°C water bath for exactly 2 minutes, then returned to ice for further 15 minutes. The sample was then warmed to room temperature, 800µl LB broth was added and tube was incubated at 37°C for 45 minutes. Aliquots of the samples (200 µl, 100µl, 50µl, 10µl) were plated out onto LB agar plates containing ampicillin. To identify bacterial colonies containing recombinant plasmid by α-complementation, plates containing X-gal and IPTG. 40µl of X-gal 20mg/ml and 4µl of IPTG 200mg/ml were spread onto pre-made plates and the solution was allowed to dry before adding bacteria suspension. The plates were incubated in inverted position overnight at 37°C.

2.2.9 Small scale preparation and purification of plasmid DNA

Minipreparations of plasmid DNA were obtained by two methods depending on the quality of DNA needed. An alkaline lysis method was used to obtain crude preparations for checking plasmid size by electrophoresis and performing restriction digests. QIAprep spin columns (Qiagen) were used to obtain high quality plasmid DNA for procedures such as sequencing, *in vitro* transcription and subcloning.

A single colony was picked up and transferred into a 25ml tube containing 5ml LB broth and ampicillin (100µl/ml). The tube was incubated overnight at 37°C with shaking.

2.2.9.1 Plasmid preparation by alikali lysis

An aliquot (1.5ml) of the overnight bacterial culture was then transferred into a

microcentrifuge tube and centrifuged at 20,000g for 2 minutes. The supernatant was removed and the cell pellet was resuspended in 100µl ice-cold solution I (50 mM glucose, 25 mM Tris-HCL pH 8.0, 10mMEDTA). After two hundred microliters of freshly prepared solution II (0.2 M NaOH, 1% SDS) were added and mixed gently, the tube was left on ice for 5 minutes. The incubation was followed by the addition of 150 µl solution III (2.9M potassium acetate, 11.5% (v/v) glacial acetic acid), the tube vortexed for 10-20 second and the cell debris was pelleted by centrifugation at 20,000g for 5 minutes. The supernatant was transferred to clean tube containing 400µl phenol:chloroform:isoamyle alcohol, vortexed briefly to mix and centrifuged for 5 minutes as before. The aqueous layer was transferred to a clean tube, 800 µl ethanol was added and left at - 20°C for 20 minutes and then centrifuged for 5 minutes as before. The pellet was washed with 70% ethanol, air dried and suspended in 50 µl distilled water containing 50 µg/ml RNase A (Sigma).

2.2.9.2 Plasmid preparation using QIAprep spin columns

1-5 ml of the overnight cultured plasmid DNA was pelleted and the supernatant was discarded. The plamid DNA was then isolated using a QIAprep spin plasmid kit (Qiagen) according to the manufacture's instructions. The preparation yielded up to 20µg of plasmid DNA in the 50-100 µl of water.

2.2.10 Large scale preparation of plasmid DNA

The Qiagen Maxiprep kit was used when large amount of high quality DNA was required. 250ml of culture cells centrifuged in a Beckman J2-21 M/E centrifuge with a JA 14 rotor at 4000 rpm for 10 minutes at 4°C. The DNA was then prepared using the reagents and instructions supplied by the manufacturer. This usually gave a yield of between 1-2mg.

A variation of the alkaline lysis mouthed of Sambrook *et al.* (1989) was used to obtain large quantities of DNA. A single colony was picked and transferred into a 25 ml tube containing 10 ml LB broth and ampicillin (100µl/ml). The tube was incubated overnight at 37°C with shaking. The next day 1ml of this overnight culture was incubated in 500ml of LB with ampicillin (100µl/ml) in a 2 litre flask at 37°C

overnight with shaking. The following day the culture was split into 250ml tube and the cells centrifuged at 4000 rpm for 15 minutes at 4°C. The pellets were reseuspended in a total of 10ml of ice cold Solution I: 50 mM glucose, 25 mM Tris-HCl, pH 8.0, 10 mM EDTA, pH8.0 and transferred to a 50ml Falco tube. When the cells had resuspended, 20ml of freshly prepared Solution II: 0.2M NaOH, 1%SDS, was added into it and the contents were mixed by gently inverting. The tubes were left on ice for 10 minutes after which 15ml of ice cold Solution III: 3M potassium acetate, 2M glacial acetic was added, shaken to mix and stored on ice for 10 minutes, then centrifuged as above. The supernatant was now removed and DNA then precipitated by adding 0.7 volumes of isopropanol. The DNA was recovered by centrifugation at 4000 rpm for 20 minutes at 4°C, washed in 70% ethanol, air dried and resuspended in 6ml of TE, pH8.0.

2.2.11 Purification of plasmid DNA by CsCl/ethidium bromide centrifugation

The plasmid DNA was purified by caesium chloride density gradient method described in Sambrook *et al.* (1989). 25g of CsCl were dissolved in 20 ml of water with 7.5 ml of the resuspended DNA in TE. Using a syringe, this solution was transferred to a Beckman Quick-Seal tube and 1.25 ml of 10mg/ml of ethidium bromide added. The tubes were topped-up with liquid paraffin, balanced and heat-sealed. They were then centrifuged at 45,000 rpm in Beckman Vti 50 fixed rotor, for at least 16 hours at 23°C.

The tubes were viewed under U.V light and the plasmid band was removed using a 5ml syringe with a 19-gauge needle. The solution was extracted 3 times with an equal volume of isopropanol to remove the ethidium bromide. It was then dialysed against 2 litres of TE for overnight in a cold room. The DNA was transferred into a tube, with 300µl LiCl (10mM) and ethanol. The DNA was precipitated at -70°C for 30 minutes and recovered by centrifugation at 8000 rpm for 30 minutes at 4°C. The DNA was washed with 70% enthanol, dried and resuspended in 0.5ml of TE.

2.2.12 Recovery and purification of DNA fractionated on agarose gels

DNA samples were electrophoresed on an agarose gel of appropriate concentration,

which would allow separation, and resolution of the DNA fragment to be excised. Electrophoresis continued until the DNA fragment of interest had separated significantly enough to be cut from the gel. The fragment of interest was then visualised using an UV light box and cut from the gel.

2.2.12.1 Qiagen

The purification of the DNA within the agarose was performed by the electrocution method described by Sambrook *et al.*, (1989). More recently this procedure was replaced by purification with QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer's instructions.

2.2.12.2 Geneclean

The Geneclean II kit (Bio 101 Inc.) was used according to the manufacturer's instructions to remove and purify DNA from agarose gels and to desalt solutions of DNA.

2.2.13 DNA sequencing by dideoxy chain-termination

DNA sequencing was performed by the dideoxy chain-termination proposed by Sanger *et al.* (1977) using Sequencing kit version 2.0 (Amersham). The DNA was prepared using QIAprep Spin Plasmid kit (Qiagen) or QIAprep Maxiprep kit (Qiagen). Approximately 2µg DNA was made up to 20µl with TE and denatured by adding 5µl of 1M NaOH, 1mM EDTA and incubating at room temperature for 5 minutes. The DNA was precipitated with 5µl 3M NH_4Ac and 75µl ethanol. The pellet was resuspended in 7µl H_2O, 1µl of primer (10µg/ml) and 2µl 5 X Sequenase buffer were added and the reaction was heated to 65°C for 3 minutes and incubated at 37°c for 30 minutes. The remaining components of the sequencing reaction were added and the sequencing was carried out according to the manufacturer's instructions.

2.2.14 Cycle sequencing with thermo Sequenase DNA polymerase

Murray (1989) described cycle sequencing as the "sequencing linear amplification

version" of the polymerase chain reaction (PCR). Double stranded DNA sequencing was carried out using the Thermo Sequenase DNA polymerase (Amersham) following the manufacturer's instructions.

This technique combines the use of thermostable polymerases and thermocycling procedure with dideoxy sequencing. The first stage was a labelling step, in which the primer was extended using three of four deoxynucleotide triphosphates including one ^{35}S-labelled dNTP. In this step the reaction mixture was thermally cycled (15 seconds at 95°C, 30 seconds at 60°C continued for 50 cycles). In the second step, the concentration of all the deoxynucleoside triphosphates was increased, a dideoxynucleoside triphophate was added, and the reaction was again thermally cycled (30 seconds at 95°C 100 seconds at 72°C) for 45 cycles. The reaction products were finally denatured by the addition of manide and heated to 70°C for 10 minutes. Analysed samples were denaturing polyacrylamid gel electrophored.

2.2.15 RNase protection assays

RNase protection assays were carried out using the technique described by Zinn *et al.* (1983). In this process, RNase protection assays kit (RPA II, Ambion) was used according to the manufacturer's instructions.

10-30 µg of total RNA was added to 2-4 X 10^4 cpm of probe, made up to 20µl with water and co-precipitated at -70°C for 30 minutes by adjusting the concentration of NH_4OAc to 0.5 M and adding 2.5 volumes of ethanol. The precipitate was recovered by centrifugation at 12,000 X g for 20 minutes at 4°C and the pellet dried. The pellet was resuspended in 20µl of hybridisation buffer: 80% deionized formamide, 100 mM sodium citrate pH 6.4, 300 mM sodium acetate pH 6.4, 1 mM EDTA, heated at 90°C for 5 minutes and hybridised overnight at 42°C.

After hybridisation was complete, 200µl of RNase digestion buffer: 10mM Tris-Cl, pH 7.5, 5 mM EDTA, 300 mM NaCl containing 5µg/ ml RNase A/20U/ml RNase T1 Mix was added and the reaction incubated at 37°C for 30 minutes. Following RNase digestion, 300µl of RNase inactivation/precipitation mix was added and precipitated at -20°C for 30 minutes. Samples were then centrifuged at 12,000 X g for 15 minutes. The pellets were resuspended in 8µl of denaturing gel-loading buffer: 90% (v/v) formamide, 10mM EDTA, pH 8.0, 0.01% (w/v) xylene cyanol, 0.01% bromophenol

blue. The samples were heated at 95°C for 5 minutes and electrophoresed on a 6% denaturing polyacrylamide gel (see Section 2.2.4) at 60 Watts until the bromophenol blue was about 1 cm from the bottom.

2.2.16 Labelling of nucleic acid probes

2.2.16.1 End-lablling of oligonuleotide with T4 polynucleotide kinase

Oligonucleotides (10 pmol/µl) were combined with 5 X Forward Reaction buffer (supplied with the T4 polynuleotide kinase), 40 µCi [γ-^{32}P] ATP and 10 units T4 polynuleotide kinase (Gibco-BRL) in a volume of 10µl and incubated at 37°C for 30 minutes. The reaction was incubated at 65°C for 10 minutes to inactivate the enzyme. The unincorporated [γ-^{32}P] ATP was removed by purification through CROMA SPIN-10 columns (Clonthch Laboratories, Inc.) with respect to the manufacturer's instructions.

2.2.16.2 Random prime labelling with Klenow I and [α-^{32}P]dCTP

Probe fragments were isolated through restriction endonulease digestion and subjected to gel was electrophoresis and extracted. Probe fragments were labelled by random prime labelling for the purpose of probe in Southern blotting, and library screening was conducted by the method of Feinberg and Vogelstein (1983).

The Oliglabelling kit (Pharmacia) was used to synthesise probes based on the manufacturer's instructions. Probes were synthesised by heating 50ng of probe fragments in a total volume of 34µl to 95°C for 3 minutes, and then the sample was placed on ice for 2 minutes. 10µl of reagent mix (provided), 5µl (50µCi) α-^{32}PdCTP and 5 units of Klenow fragment of DNA polymerase I were added and incubated at 37°C for 60 minutes. Unincorporated nucleotides were removed using MicroSpin S-300 columns (Pharmacia) according to the manufacturer's instructions. The purified probe was then heated at 100°C for 5 minutes, cooled on ice and added to the relevant hybridisation buffer.

2.2.16.3 *In vitro* transcription of [α-^{32}P] labelled RNA

High specific activity RNA probes were prepared by *in vitro* transcription in the presence of [α-^{32}P]CTP for use in RNase protection assays.

The pBluescript II vector used in this project contained promoters specific for T3 and T7 RNA polymerase flanking polylinker site. The plasmid DNA of interest was linearized by restriction endonuclease digestion of a chosen site within the polylinker. This linearized plasmid was then used for the synthesis of "run off" sense and antisense probes of the insert. An appropriate restriction endonuclease was used to linearize 2µl of plasmid DNA. The plasmid was electrophoresed on a 1.0% agarose gel and purified from the gel using Qiagen. The *In vitro* transcription kit (Ambion) was then used to synthesise sense and antisense probes. Plamids were transcribed in a 20µl reaction containing 1 X transcription buffer, 10µM CTP, 0.5 mM NTP (ATP, UTP, GTP), 50µCi [α-^{32}P]CTP, 500ng of template (linearized DNA) and 20 unite of T7 or T3 RNA polymerase and the reaction was incubated at 37°C for 60 minutes. 2 units of RNase-free DNase I were added and incubated at 37°C for 15 minutes. These were subjected to gel electrophoresis as described in Section 2.2.23 to determine the efficiency of the reaction. The gel plates were then removed from the tank and the gel was covered in Saran Wrap and exposed to X-ray for 1-2 minutes. The autoradiograph was then placed on top of the gel and the area representing the probe was cut from the gel using a razor blade. The gel slice was diced, placed in a microcentrifuge tube and then 350 µl of Probe Elution Buffer (Ambion) was added, after all, the RNA was left to elute 2-4 hours or overnight at 37°C.

2.2.17 Southern blotting

For plasmid southern blots, the plasmid or restriction digest and 0.5µg of 1Kb marker (Gibco BRL) were run on a 1.0 % nondenaturing agarose gel, at 25 volts overnight. The gel was stained in 0.5µg/ml ethium bromide in distilled water for 15 minutes for photograph. Southern blotting was then carried out using the method of Southern (1975). To increase the efficiency of transfer of high molecular weight DNA the gel was first treated with 0.25M HCl for 20 minutes at room temperature. Then the gel was the rinsed in distilled water and treated with denaturation buffer; 1.5M NaCl,

0.5M NaOH for 30 minutes. The gel was further rinsed in distilled water and placed in neutralisation buffer containing 0.5M Tris-HCl, 1.5M NaCl, pH 8.0, for 30 minutes.

A piece of filter paper (3MM Whatman) moistended with 20 x SSC was placed on top of a platform in tray of 20 x SSC so that ends of the filter paper were submerged in the 20 x SSC. The product obtained above was placed on top of the filter paper and the exposed filter paper covered with cling film. A piece of nylon membrane (Hybond-N, Amersham) was cut with a 1mm border over the gel, the filter was moistened in distilled water then immersed in 20 x SSC for 10 minutes. The filter was then placed onto the surface of the gel, which must not have air bubbles formed. Two pieces of filter paper were cut to the same size as the gel were soaked in 2 x SSC and placed onto the surface of the nylon membrane. Tissue paper was stacked on top of this and weighed down with a water-filled 250 ml bottle weight. The gel was left to blot for overnight. The nylon filter was rinsed in 2 x SSC for 15 minutes and the DNA were fixed to the nylon filter by baking at 80°C under vacuum for 2.0 hours or UV-crosslinked (Stratagene) by laying them face down on a UV transilluminator for 1 minutes.

Hybridisation and washing procedures were carried out in Hybaid C4 glass tubes. The membrane was prehybridised in QuikHyb hybridisation solution (Stratagene) at 68°C for 60 minutes. The double-stranded probe (see in Section 2.2.15.2) and salmon sperm DNA mixture were boiled for 3 minutes, and then added the mixture to the pre-hybridisation solution in the glass tubes, and hybridised at 68°C for 4 hours.

Filters were washed twice for 15 minutes at room temperature with a 2 x SSC, 0.1% (w/v) SDS wash solution and then washed once for 30 minutes at 60°C with 0.1 x SSC, 0.1% (w/v) SDS wash solution for a high-stringency wash. The filters were blotted on filter paper and while still damp, were wrapped in Saran Wrap and exposed to X-ray film with an intensifying screen at -70°C.

2.2.18 Polymerase chain reaction

PCR was used for several different purposes: identification of the upstream region of the CRH-R2β gene, construction of plasmid for promoter analysis, RT-PCR was used to identify the CRH-R subtypes in human myometrial tissue and 5'RACE PCR determination of CRH-R2 transcription initiation site or 5'-cDNA of CRH-R1. PCR

was also used to generate probes fragments for hybridisation to Southern filters and to analyse cultures or bacterial colonies for the presence and size of inserts.

2.2.18.1 RT-PCR

2.2.18.1.1 Reverse transcription

Total RNA (3--5μg) was heated at 75°C for 5 minutes and cooled on ice, reverse transcription in 20μl reaction was carried for 1 hour at 42°C. The reaction contains 200-400ng Oligo$(dT)_{18}$ or 5 pmoles of a gene specific primer, 0.5mM each of four dNTPs (Promega), 50mMTris- HCl pH 8.3, 3mM $MgCl_2$ 10 mM DTT and 200 units Reverse Transcriptase (Gibco-BRL). The reaction was stopped by incubation at 70°C for 15 minutes. The RNA was then removed by digestion with RNase H (2 units), and the incubation continued for further 20 minutes at 37°C.

2.2.18.1.2 PCR

To amplify CRH-R subtype cDNA, only 10% of the reverse transcription reactions were used for PCR. cDNA was amplified by PCR in 50μl reaction containing 20 mM Tris- HCl (PH 8.4), 50mM KCl, 1.5mM $MgCl_2$, 200μM dNTPs and 2.5 units *Taq* DNA polymerase (Gibco-BRL). 10μM each, of forward and reverse primers (see in Section 2.1.8) were also included.

For the first round PCR, the sense primer was used in combination with an antisense primer adapter-(dT) with nucleotide sequence 5′-GTCGACAAGC$(T)_{18}$-3′. The products of this reaction were served as templates for the second round of amplification, which two internal primers were adopted. (1) First round PCR: after an initial denaturation at 95°C for 2 minutes, cooled to 72C° at which point 2.5 units *Taq* DNA polymerase (Gibco-BRL) was added. 40 cycles of PCR were carried out, denaturation at 95°C for 45 seconds, annealing at 55°C for 45 seconds, extension at 72°C for 1.5 minutes. The final extension was at 72°C for 15 minutes. (2) Second round PCR: after an initial denaturation at 95°C for 2 minutes, cooled to 72°C, at which point 2.5 units *Taq* DNA polymerase (Gibco-BRL) was added. 40 cycles of PCR were carried out, denaturation at 95°C for 45 seconds, annealing at 58°C for 45

seconds, extension at 72°C for 1.5 minutes. Final extension was at 72°C for 15 minutes. As a negative control for all the reactions, distilled water was used in place of the cDNA.

The reactions were analysed by agarose gel electrophoresis on a 1.0% agarose gel as described in Section 2.2.2.1 and the amplified products were gel isolated as described in Section 2.2.11.

2.2.18.2 Genomic walking

Approximately 1-2 µg of human genomic DNA (from peripheral blood leucocytes) was digested to completion with 80 units of *Nco*I (GibcoBRL). The sample was extracted with an equal volume of phenol:chloroform:isoamyl alcohol and precipitated with the addition of 2.5 volumes 100% ethanol. The pellet was washed with 70% ethanol and resuspended in 10µl of sterile distilled H_2O. The DNA was boiled for 5 minutes then quickly cooled on ice. The entire sample was incubated with 0.5mM dATP and 1.5mM $CoCl_2$ in 20µl of terminal transferase (TdT) buffer containing 50 units of TdT (Amersham) at 37°C for 60 minutes. The reaction was stopped by heating the sample at 65°C for 15 minutes. The polymerase chain reaction was performed using 2µl (100ng) of the polyadenylated genomic DNA as a template, 200ng of gene-specific primer

5'-GAGGAGGTGTGGGACGTAGAGGAG-3'

200ng of universal-T17 primer

5'-GAGGACTCGAGCTCAAGCTTTTTTTTTTTTTTTT-3',

200uM dNTPs and 1ul of Advantage PCR 2 polymerase Mix (Clontech) in 50ul volume. The PCR was carried out at 95°C for 1 minute, then 95°C for 30 seconds, 72°C for 1 minute, 5 cycles, 95°C for 30 seconds, 70°C for 1 minute, 5 cycles, 95°C for 30 seconds, 68°C for 1 minute cycled 35 times. Following the reaction, 1µl of the sample was then used as a template for the second round of PCR under the same condition, except that 200ng of gene- specific primer

5'-GGTATTCTGGACCCACTTCTTCCAC-3'

and 200ng of universal primer

5'-GAGGACTCGAGCTCAAGC-3'

were used. The final PCR product was subcloned into the pGEM-T vector (Promega).

The insert DNA was cycle-sequenced using an ABI 373A (Applied Biosystems) automated fluorescent sequencer.

2.2.18.3 5'-Rapid amplification of cDNA end (5'RACE)

2.2.18.3.1 5'RACE to identify the 5'-cDNA of CRH-R1

In order to confirm the sequence of the 5′-cDNA of CRH-R1α, the 5′-RACE system Version 2.0 (Gibco-BRL) was employed to amplify the 5′-cDNA for sequencing. The 5′-RACE system Version 2.0 kit was used according to the manufacture′s protocol. Briefly, 2.5pmol of a gene-specific primer

CRFII 5′-CCCTTCGAAGTCTGACGACACCTG-3′

was hybridised to 5 μg of human myometrial total RNA, and cDNA synthesised using Superscript II Reverse Transcriptase (Gibco-BRL). The RNA was then removed by digestion with RNase H, and the cDNA was isolated. A poly(dC) tail was then added to the 3′-terminus of the purified cDNA using dCTP and the TdT enzyme. The cDNA region corresponding to the 5′-cDNA was then amplified by two successive rounds of PCR using additional gene-specific antisense primers

CRF4 5'-CCTTCCTCAACTTCATCTAC-3'

and CRF7 5'-TACTTGTAGGTAAACGCGTCG-3',

together with the anchor primer supplied by a manufacturer. The amplified products yielded through the second round of PCR, were analysed by electrophoresis on 1.0% agarose gel. The first round PCR conditions were 40 cycles of 95°C for 45 seconds, 55°C for 45 seconds and 72°C for 1 minute followed by 72°C for 15 minutes. Aliquots of the above reaction were re-amplified with a nested, gene-specific primer CRF7 and anchor primer. PCR conditions for 40 cycles were as follow: 95°C, 45 seconds; 58°C, 45 seconds; 72°C, 60 seconds.

2.2.18.3.2 Determinations of the transcription start site.

The 5'end of the human CRH-R2β receptor cDNA was obtained by the rapid amplification of cDNA 5'-ends(5'RACE) procedure using the commercially prepared Marathon Ready cDNA system (Clontech) with respect to the manufacturer's

recommendations. An uncloned library of the adaptor-ligated cDNA prepared from human hippocampus was used to amplify the 5'-end of the CRH-R2β cDNA. The first round PCR was performed with a CRH-R2β cDNA specific primer (see Section 2.1.8) and plus anchor primer AP1 supplied by Clontech. A nested PCR was then performed with the Clontech plus primer AP2 and a second cDNA specific primer. The resultant products were cloned into the pGEM-T cloning vector and the clones sequenced. In order to confirm the 5'-end cDNA region, a new primer was designed based on the sequences of the above clones. The products of those fragments amplified with the new primer were characterised by cloning, sequencing, and comparison with the above sequence results.

5'RACE was used with combinations of CRH-R2α and 2γ (exon-1 gene specific primer) to clone the 5'-end cDNA from human hippocampus (Marathon-Ready cDNA, Clontech). The CRH-R2α, 2γ of 5'RACE derived from cDNA clones include the 5'-Untranslated region (UTR) upstream of the CRH-R2β open reading frame.

2.2.18.4 Promoter construct preparation

The PCR fragments spanning different regions of promoter were cloned upstream of the CAT reporter gene in the pCAT 3-Basic vector (Promega). The oligonucleotides used for this PCR have a restriction site 5' to facilitate cloning into the pCAT 3-Basic vector polylinker. For construction derived from promoter of CRH-R2, a common 3' reverse oligonucleotide

5'-CGG**CTCGAG**AGGTGTGGGACGTAGAGGAGG-3'

was used in combination with the different 5' oligonucleotides:

5'-CC**GGTACC**GACCTGTGAGATTTCACAGGAG-3' for construct -1393,

5'-CC**GGTACC**GAGGCTCTGGAAGAACAGACC-3' for construct -1035,

5'-CC**GGTACC**ACCTACCCACAGTACAGTATC-3' for construct –553.

The PCR fragments were digested with *Kpn* I and *Xho* I and cloned into equivalent site of the pCAT3-basic vector.

2.2.18.5 PCR for analysing colonies for the presence and size and an insert

To analyse the clones containing the exon-1 of CRH-R1α gene, 25ng of minipreped DNA was used in a PCR reaction but only 20 cycles not 30 cycles were carried out. The gene specific primers PE5 and PE3 correspond to +2 to +23 nt and +239 to 260 nt in exon-1. The reaction was then analysed by electrophoreses on a 1.0% agarose gel. Colonies were picked up using a sterile tip and suspended in 100µl distilled TE. The tip was then used to seed a 50µl starter culture of L Broth with ampicillin. The suspension containing the colony was then heated to 95°C for 10 minutes, centrifuged at 13,000 rpm for 5 minutes in a bench top centrifuge and 5µl of the "boilate" was used in the PCR. Each PCR reaction contained 1X PCR buffer (Gibco-BRL), 1.5mM $MgCl_2$, 0.05% W-1, 0.25mM dATP, 0.25mM dCTP, 0.25mM dGTP, 0.25mM dTTP, 2.5µM of each primer and 2.5units *Taq* polymerase (Gibco-BRL). The primer used in the reaction depended on the plasmid to be analysed. The primers T7 and SP6 were used for pGEM-T vector, and T3 and T7 primer were used for pBluescript.

2.2.18.6 Genomic PCR for amplify the region between the CRH-R2β and 2γ exon

In order to obtain the genomic sequence between the CRH-R2β and CRH-R2γ exons, PCR was performed with Elongase system (Gibco BRL) and human genomic DNA from peripheral blood leucocytes. (1) PCR reaction, we used a primer to the CRH-R2β exon 5'-end (B2: 5'-CTGGCATGAGGGGTCCCTCAG-3') and a primer to the 5'-end of the CRH-R2γ specific exon (G3: 5'-GTAGGCAGATTGCTTGAGAC-3'). (2) PCR reaction, we used a primer to the CRH-R2β exon 3'-end (B4: 5'-CATGACCCTCACCAACCTCTC-3') and a primer to the 5'-end of the CRH-R2γ specific exon (G3: 5'-GTAGGCAGATTGCTTGAGAC-3'). PCR was performed with 500ng of DNA in the presence of each primer at 200 nM, each dNTP at 200µM, 60 mM tris/sulphate, pH 9.1, 18 $mM(NH_4)_2SO_4$, 1.6mM $MgSO_4$ and 1µl of Elongase enzyme mix. Amplification was performed for 35-40 cycles with denaturation at 94°C for 30 second followed by annealing/extension at 68°C for 3-10 minutes. The PCR products were analysed by 1% agarose-gel electrophoresis in the presence of TBE buffer and then subcloned into the pGEM-T vector and sequenced.

2.2.19 Library screening

2.2.19.1 Filter preparation

The protocol was based on the standard protocol of Sambrook *et al.* (1989). Colonies were grown on a "master" filter and one replica was made for screening. The "master" filter was stored at 4°C for recovery of positive plasmids. Approximately 3 x 10^6 cosmid containing the transformed library (human placental genomic cosmid DNA library, Clontech Laboratories, Inc) were plated on 4 x (20 x 20cm) Hybond N (Amersham) membrane placed on 4 x (22 x 22cm) LB-agar which contained 50µg/ml ampicillin and grown at 37°C until the colonies were approximately 1mm in diameter. This master filter was then placed on a sterile 3MM Whatmann paper on a flat glass surface. A fresh 20 x 20 cm Hybond N membrane was placed on a fresh LB-agar which contained 50µg/ml ampicillin plate to moisten it, and was then placed on the master filter. Another sterile 3MM Whatmann paper was then placed on top. The second glass plate was placed on the stack and pressure applied too firmly press the plates together. The upper plate and 3MM Whatmann paper were then removed and the orientation of the filter was labelled using a small needle. The master and replica filters were separated and placed face up on LB-agar/ampicillin plates. The master filters were incubated at 37°C for 4 hours and sealed with cling film and then stored at 4°C. The replica filters were incubated at 37°C for 6-8 hours until the colonies were approximately 1 mm in diameter.

2.2.19.2 Preparation of replica filter

The filter was placed colony-side up on the surface of 3MM Whatmann filter paper previously prewetted with 0.5 M NaOH and left for 3 minutes. Then the filter was placed on another set of 3MM Whatmann filter paper previously pre-wetted with 1M Tri-HCl (pH 7.5) and left for 3 minutes. Afterward, the filter was placed on the third set of 3MM Whatmann filter paper previously pre-wetted with 1 M Tris-HCl (pH7.5) and 1.5 M NaCl and left for 3 minutes. After all above has been done, the filter was removed into a solution of 1 M Tris-HCl (pH7.5) and 1.5 M NaCl. With a gloved hand, the bacterial debris was removed by gently rubbing the filter paper. 3MM

Whatmann filter paper was used to blot off excess liquid, and the filter was air-dried. The filter was baked for 1 hour at 80°C in a vacuum oven or the DNA UV-crosslinked to the filter by laying it DNA-side down on an UV transilluminator for 1 minute.

Hybridisation and washing procedures were carried out in Hybaid C4 glass tubes. The membrane was pre-hybridised in QuikHyb hybridisation solution (Stratagene) at 68°C for 60 minutes. The double-stranded probe (see in Section 2.2.15.2) and salmon sperm DNA mixture were boiled for 3 minutes, and then added the mixture to the pre-hybridisation solution in glass tubes, and hybridised at 68°C for 4 hours.

Filters were washed twice for 15 minutes at room temperature with a 2 x SSC, 0.1% (w/v) SDS wash solution and then washed once for 30 minutes at 60°C with 0.1 x SSC, 0.1% (w/v) SDS wash solution for a high-stringency wash. The filters blotted on filter papers while they still damp, were wrapped in Saran Wrap and exposed to X-ray film with an intensifying screen at -70°C for 3-5 days.

2.2.19.3 Second round screening

The position of positive colonies on the master filter was determined and marked using the mm square plastic sheets with labelled marker holes. A 0.5 x 0.5 cm region surrounding the desired clone was excised. The filter piece was placed in a 25 ml tube containing 5 ml LB broth and ampicillin (100µl/ml). The tube was incubated overnight at 37°C with shaking. 10µl of this suspension was diluted in 1ml LB broth. This dilution was then further diluted to 1:10, 1:15, 1:20 and 1:100 with LB broth which contained ampicillin (100µl/ml), and 100µl of each dilution was plated on LB-agar which contained ampicillin (100µl/ml) and incubated at 37°C overnight.

Plates with 400-600 colonies were screened. Hybond N (Amersham) filters were cut to sizes and placed on the colonies and their orientation was marked by 3 holes formed using a small needle. The filters were treated as described in the first round screening. The plates were incubated for 2 hours and then stored at 4°C. The baked, crosslinked filters were hybridised and washed as described for the filters in the first round screening and exposed to X-ray film in a cassette with an intensifying screen for several days at -70°C.

2.2.19.4 Third round screening

Positive colonies were picked and resupended in 5ml LB broth. Dilutions and screening were carried out as for the second round screening. Filters were treated, hybridised and washed as described above.

Positive colonies from the third round screening were picked, minipreped using QIAprep spin plasmid kit (Qiagen), digested with restriction endonucleases, electrophoresed on a 1.0% agarose gel. The gel was Southern blotted as described in Section 2.2.16 and the plasmid Southern probed to ensure that the clones contained positive inserts.

2.2.20 Transfection of eukaryotic cells

Cell lines employed in reporter gene studies were transiently transfected with plasmid expression vectors.

Confluent cell flasks were trypsinized, the number of cells counted and in six-well plates (Nunclon) were seeded with 3 x 10^5 cells per well. Cells were grown in 2 ml of Dulbecos modified Eagle medium, 10% fatal bovine serum, penicillin (200Units/ml), streptomycin (200mg/ml) and incubated in 5% CO_2 at 37°C for 24h before transfection. All DNA plasmid used in transfection were purified using QIAGEN plasmid Maxi Kit (Qiagen) and were analysed by agarose gel electrophoresis to verify plasmid integrity. The pCAT3-basic-less plasmid was routinely in all transfection experiments as negative controls.

The cells were transfected using Lipofectamine Reagent (Gibco-BRL) in 1ml of serum-free medium (Optimem) for 3 hours according to the manufacture's instruction. Medium was replaced with 2ml of fresh DMEM containing 10% fetal bovine serum 24h after transfection. After the start of transfection of 48 hours, the cells were harvested.

2.2.21 Analysis of extract from transfected cells

2.2.21.1 preparation of cell extracts

The wells containing monolayers of transiently transfected cells were washed three

times with 3ml of ice cold PBS. After last work step, the cells were removed PBS with a fine-tipped Pasteur pipette and then added 340µl of lysis buffer (Boehringer mannheim) and allowed to stand for 30 minutes at room temperature. The cells were transferred into microcentrifuge tubes and centrifuged at 13,000 rpm for 10 minutes at 4°C. The supernatants were transferred into microcentrifuge tubes. Cell lysis was ensured by three cycles of freeze-thawing (2 minutes in ethanol/dry ice then 1 minute in 37°C water bath). The cell debris was subsequently removed by centrifugation at 13,000 rpm for 10 minutes. The cell extracts were finally transferred into clean tubes and stored at - 70°C until use.

2.2.21.2 Chloramphenicol acetyltransferase (CAT) assay

The CAT activity of cytoplasmic extracts from cells transiently transfected with CAT expression vectors was measured by an enzyme-linked immunosorbent assay (ELISA) method.

The CAT activity was measured using a CAT ELISA kit (Boehringer, Mannheim) according to the manufacturer's instruction. The cells extracted were added with 10µl into the well (10µl/well) instead of the recommend 200µl/well.

Chapter 3 Amplifying and cloning CRH-R1α, 1β, and 1c from human myometrium

3.1 Introduction

The aim of the first part of the project was to identify CRH-R subtype expression in human myometrial biopsies (pregnant and non-pregnant) and myometrial cell cultures. PCR primers were designed specifically for the CRH-R subtypes and their presence was confirmed by using RT-PCR and nucleotide sequencing. RNase protection assays was also employed to study CRH-R mRNA expression in the human myometrium.

3.2 General description of PCR and CRH-R1

Since protein synthesis is nearly always directly proportional to the level of mRNA encoding the peptide, the initial experiments on gene regulation are usually directed to quantitating changes in specific mRNAs. For example, the techniques of *in situ* hybridisation and the reverse transcriptase polymerase chain reaction (RT-PCR) can be used to localise the expression of specific mRNAs to individual cells or tissues. Other methods such as RNase protection have allowed detection of and discrimination between variant forms of similar mRNAs.

The polymerase chain reaction (PCR) is a method which entails the use of two primers. The two primers hybridise to opposite strands of a template and are orientated so that there are 3'ends pointing towards each other. These primers are then used in an amplification reaction in which the product is the region between the two primers. The DNA segment amplified in the first round is then used as a template in subsequent cycles with the products accumulating in an exponential fashion (Saiki *et al.*, 1988). PCR is a powerful technique for amplification of DNA and is the most sensitive technique available to measure extremely low levels of mRNA. The first step in measuring mRNA is to use the isolated RNA as a template to synthesise

complementary DNA with the viral enzyme reverse transcriptase. It is now possible to obtain accurate and sensitive measures of gene expression in small samples of tissue, or in a few cells *in vitro*, using RT-PCR. This represents a considerable increase in sensitivity over other analytical methods such as northern blot analysis or RNase protection assays.

The CRH-R1α receptor has been cloned from several species including human (Chen *et al.*, 1993), mouse (Vita *et al.*, 1993), chicken (Yu *et al.*, 1996), and rat (Perrin *et al.*, 1993, Chang *et al.*, 1993). Species homologues are 98% identical over their full length of 415 amino acids. In general, the CRH-R1 shares approximately 30% identity with all other members of the neuropeptide receptor family. The CRH-R1α is a 415-amino acid protein comprising seven putative membrane-spanning domains characteristic of most G protein-coupled receptors. The CRH-R1 has three splice variants: 1 β (444 AA), 1c (375 AA) and 1d (401 AA). The CRH-R1α encodes 29 additional amino acids inserted in the first intracellular loop, named: CRH-R1β, and CRH-R1c has a 40 amino acids deletion in the N-terminal extracellular domain. The structure CRH-R1α, 1β, 1c and 1d is shown in Figure 3.1.

The CRH-R1

CHR-R1α CCCGGGCATTCAGGACGGTAGCCGAGCGAGCCCGAGGATGGGAGGGCACCCGCAGCTCC(59)
1β ATGGGAGGGCACCCGCAGCTCC
1c AGCCGAGCGAGCCCGAGGATGGGAGGGCACCCGCAGCTCC
1d ATGGGAGGGCACCCGCAGCTCC
↓1
GTCTCGTCAA GGCCCTTCTC CTTCTGGGGC TGAACCCCGT CTCTGCCTCC CTCCAGGACC AGCACTGCGA (129)
GTCTCGTCAA GGCCCTTCTC CTTCTGGGGC TGAACCCCGT CTCTGCCTCC CTCCAGGACC AGCACTGCGA
GTCTCGTCAA GGCCCTTCTC CTTCTGGGGC TGAACCCCGT CTCTGCCTCC CTCCAGGACC AGCACTGCGA
GTCTCGTCAA GGCCCTTCTC CTTCTGGGGC TGAACCCCGT CTCTGCCTCC CTCCAGGACC AGCACTGCGA
↓2
GAGCCTGTCC CTGGCCAGCA ACATCTCAGG ACTGCAGTGC AACGCATCCG TGGACCTCAT TGGCACCTGC (199)
GAGCCTGTCC CTGGCCAGCA ACATCTCAGG ACTGCAGTGC AACGCATCCG TGGACCTCAT TGGCACCTGC
GAGCCTGTCC CTGGCCAGCA ACATCTCAG ....
GAGCCTGTCC CTGGCCAGCA ACATCTCAGG ACTGCAGTGC AACGCATCCG TGGACCTCAT TGGCACCTGC

TGGCCCCGCA GCCCTGCGGG GCAGCTAGTG GTTCGGCCCT GCCCTGCCTT TTTCTATGGT GTCCGCTACA (269)
TGGCCCCGCA GCCCTGCGGG GCAGCTAGTG GTTCGGCCCT GCCCTGCCTT TTTCTATGGT GTCCGCTACA
....
TGGCCCCGCA GCCCTGCGGG GCAGCTAGTG GTTCGGCCCT GCCCTGCCTT TTTCTATGGT GTCCGCTACA
↓3
ATACCACAAACAATGGCTAC CGGGAGTGCC TGGCCAATGG CAGCTGGGCC GCCCGCGTGA ATTACTCCGA (339)
ATACCACAAACAATGGCTAC CGGGAGTGCC TGGCCAATGG CAGCTGGGCC GCCCGCGTGA ATTACTCCGA
..... ACAATGGCTAC CGGGAGTGCC TGGCCAATGG CAGCTGGGCC GCCCGCGTGA ATTACTCCGA
ATACCACAAACAATGGCTAC CGGGAGTGCC TGGCCAATGG CAGCTGGGCC GCCCGCGTGA ATTACTCCGA
↓4
GTGCCAGGAGATCCTCAATG AGGAGAAAAA AAGCAAGGTG CACTACCATG TCGCAGTCATCATCAACTAC (409)
GTGCCAGGAGATCCTCAATG AGGAGAAAAA AAGCAAGGTG CACTACCATG TCGCAGTCAT CATCAACTAC
GTGCCAGGAGATCCTCAATG AGGAGAAAAA AAGCAAGGTG CACTACCATG TCGCAGTCAT CATCAACTAC
GTGCCAGGAGATCCTCAATG AGGAGAAAAA AAGCAAGGTG CACTACCATG TCGCAGTCAT CATCAACTAC
↓5
CTGGGCCACT GTATCTCCCT GGTGGCCCTC CTGGTGGCCT TTGTCCTCTT TCTGCGGCTC AG..... (471)

CTGGGCCACT GTATCTCCCT GGTGGCCCTC CTGGTGGCCT TTGTCCTCTT TCTGCGGCTC AGGCCAGGCT
CTGGGCCACT GTATCTCCCT GGTGGCCCTC CTGGTGGCCT TTGTCCTCTT TCTGCGGCTC AG.....
CTGGGCCACT GTATCTCCCT GGTGGCCCTC CTGGTGGCCT TTGTCCTCTT TCTGCGGCTC AG.....

....

GCACCCATTG GGGTGACCAG GCAGATGGAG CCCTGGAGGT GGGGGCTCCA TGGAGTGGTG CCCATTTCA
....
....

..... G AGCATCCGGT GCCTGCGAAA CATCATCCAC TGGAACCTCA TCTCCGCCTT CATCCTGCGC (532)
GGTTCGAAGG AGCATCCGGT GCCTGCGAAA CATCATCCAC TGGAACCTCA TCTCCGCCTT CATCCTGCGC
..... G AGCATCCGGT GCCTGCGAAA CATCATCCAC TGGAACCTCA TCTCCGCCTT CATCCTGCGC
.....G AGCATCCGGT GCCTGCGAAA CATCATCCAC TGGAACCTCA TCTCCGCCTT CATCCTGCGC
↓6
AACGCCACCT GGTTCGTGGT CCAGCTAACC ATGAGCCCCG AGGTCCACCA GAGCAACGTG GGCTGGTGCA(602)
AACGCCACCT GGTTCGTGGT CCAGCTAACC ATGAGCCCCG AGGTCCACCA GAGCAACGTG GGCTGGTGCA
AACGCCACCT GGTTCGTGGT CCAGCTAACC ATGAGCCCCG AGGTCCACCA GAGCAACGTG GGCTGGTGCA

GGTTGGTGAC AGCCGCCTAC AACTACTTCC ATGTGACCAA CTTCTTCTGG ATGTTCGGCG AGGGCTGCTA (672)
GGTTGGTGAC AGCCGCCTAC AACTACTTCC ATGTGACCAA CTTCTTCTGG ATGTTCGGCG AGGGCTGCTA
GGTTGGTGAC AGCCGCCTAC AACTACTTCC ATGTGACCAA CTTCTTCTGG ATGTTCGGCG AGGGCTGCTA
GGTTGGTGAC AGCCGCCTAC AACTACTTCC ATGTGACCAA CTTCTTCTGG ATGTTCGGCG AGGGCTGCTA

CCTGCACACA GCCATCGTGC TCACCTACTC CACTGACCGG CTGCGCAAAT GGATGTTCAT CTGCATTGGC (742)
CCTGCACACA GCCATCGTGC TCACCTACTC CACTGACCGG CTGCGCAAAT GGATGTTCAT CTGCATTGGC
CCTGCACACA GCCATCGTGC TCACCTACTC CACTGACCGG CTGCGCAAAT GGATGTTCAT CTGCATTGGC
CCTGCACACA GCCATCGTGC TCACCTACTC CACTGACCGG CTGCGCAAAT GGATGTTCAT CTGCATTGGC
↓7 ↓8
TGGGGTGTGC CCTTCCCCAT CATTGTGGCC TGGGCCATTG GGAAGCTGTA CTACGACAAT GAGAAGTGCT (812)
TGGGGTGTGC CCTTCCCCAT CATTGTGGCC TGGGCCATTG GGAAGCTGTA CTACGACAAT GAGAAGTGCT
TGGGGTGTGC CCTTCCCCAT CATTGTGGCC TGGGCCATTG GGAAGCTGTA CTACGACAAT GAGAAGTGCT
TGGGGTGTGC CCTTCCCCAT CATTGTGGCC TGGGCCATTG GGAAGCTGTA CTACGACAAT GAGAAGTGCT
↓9
GGTTTGGCAA AAGGCCTGGG GTGTACACCG ACTACATCTA CCAGGGCCCC ATGATCCTGG TCCTGCTGAT (882)
GGTTTGGCAA AAGGCCTGGG GTGTACACCG ACTACATCTA CCAGGGCCCC ATGATCCTGG TCCTGCTGAT
GGTTTGGCAA AAGGCCTGGG GTGTACACCG ACTACATCTA CCAGGGCCCC ATGATCCTGG TCCTGCTGAT

CAATTTCATC TTCCTTTTCA ACATCGTCCG CATCCTCATG ACCAAGCTCC GGGCATCCAC CACGTCTGAG (952)
CAATTTCATC TTCCTTTTCA ACATCGTCCG CATCCTCATG ACCAAGCTCC GGGCATCCAC CACGTCTGAG
CAATTTCATC TTCCTTTTCA ACATCGTCCG CATCCTCATG ACCAAGCTCC GGGCATCCAC CACGTCTGAG
CAATTTCATC TTCCTTTTCA ACATCGTCCG CATCCTCATG ACCAAGCTCC GGGCATCCAC CACGTCTGAG
↓10
ACCATTCAGT ACAGGAAGGCTGTGAAAGCC ACTCTGGTGC TGCTGCCCCT CCTGGGCATC ACCTACATGC (1022)
ACCATTCAGT ACAGGAAGGCTGTGAAAGCC ACTCTGGTGC TGCTGCCCCT CCTGGGCATC ACCTACATGC
ACCATTCAGT ACAGGAAGGCTGTGAAAGCC ACTCTGGTGC TGCTGCCCCT CCTGGGCATC ACCTACATGC
ACCATTCAGT ACAGGAAGGCTGTGAAAGCC ACTCTGGTGC TGCTGCCCCT CCTGGGCATC ACCTACATGC

TGTTCTTCGT CAATCCCGGG GAGGATGAGG TCTCCCGGGT CGTCTTCATC TACTTCAACT CCTTCCTGGA (1092)
TGTTCTTCGT CAATCCCGGG GAGGATGAGG TCTCCCGGGT CGTCTTCATC TACTTCAACT CCTTCCTGGA
TGTTCTTCGT CAATCCCGGG GAGGATGAGG TCTCCCGGGT CGTCTTCATC TACTTCAACT CCTTCCTGGA
TGTTCTTCGT CAATCCCGGG GAGGATGAGG TCTCCCGGGT CGTCTTCATC TACTTCAACT CCTTCCTGGA
↓11 ↓12
ATCCTTCCAG GGCTTCTTTG TGTCTGTGTT CTACTGTTTC CTCAATAGTG AGGTCCGTTC TGCCATCCGG (1162)
ATCCTTCCAG GGCTTCTTTG TGTCTGTGTT CTACTGTTTC CTCAATAGTG AGGTCCGTTC TGCCATCCGG
ATCCTTCCAG GGCTTCTTTG TGTCTGTGTT CTACTGTTTC CTCAATAGTG AGGTCCGTTC TGCCATCCGG
ATCCTTCCAG ....GTCCGTTC TGCCATCCGG

AAGAGGTGGCACCGGTGGCAGGACAAGCAC TCGATCCGTG CCCGAGTGGC CCGTGCCATG TCCATCCCCA(1232)
AAGAGGTGGCACCGGTGGCAGGACAAGCAC TCGATCCGTG CCCGAGTGGC CCGTGCCATG TCCATCCCCA
AAGAGGTGGCACCGGTGGCAGGACAAGCAC TCGATCCGTG CCCGAGTGGC CCGTGCCATG TCCATCCCCA
AAGAGGTGGCACCGGTGGCAGGACAAGCAC TCGATCCGTG CCCGAGTGGC CCGTGCCATG TCCATCCCCA
—— (1285)
CCTCCCCAAC CCGTGTCAGC TTTCACAGCA TCAAGCAGTC CACAGCAGTC TGA
CCTCCCCAAC CCGTGTCAGC TTTCACAGCA TCAAGCAGTC CACAGCAGTC TGA
CCTCCCCAAC CCGTGTCAGC TTTCACAGCA TCAAGCAGTC CACAGCAGTC TGA
CCTCCCCAAC CCGTGTCAGC TTTCACAGCA TCAAGCAGTC CACAGCAGTC TGA

Figure 3.1 Comparison of the sequence of human CRH-R1 subtypes

(↓) indicates intron positions (numbers according to human CRH-R1α sequence). The initiation and termination cordons are ***underlined*** and ***overlined***, respectively. 1α, human CRH-R1α (accession number L23332); 1β, human CRH-R1β (accession number L23333); 1c, human CRH-R1c (accession number U16273) and 1d, human CRH-R1d (accession number AF 180301).

The procedures used to clone and sequence CRH-R subtypes from human myometrium are described in the following illustrate (Figure 3.2).

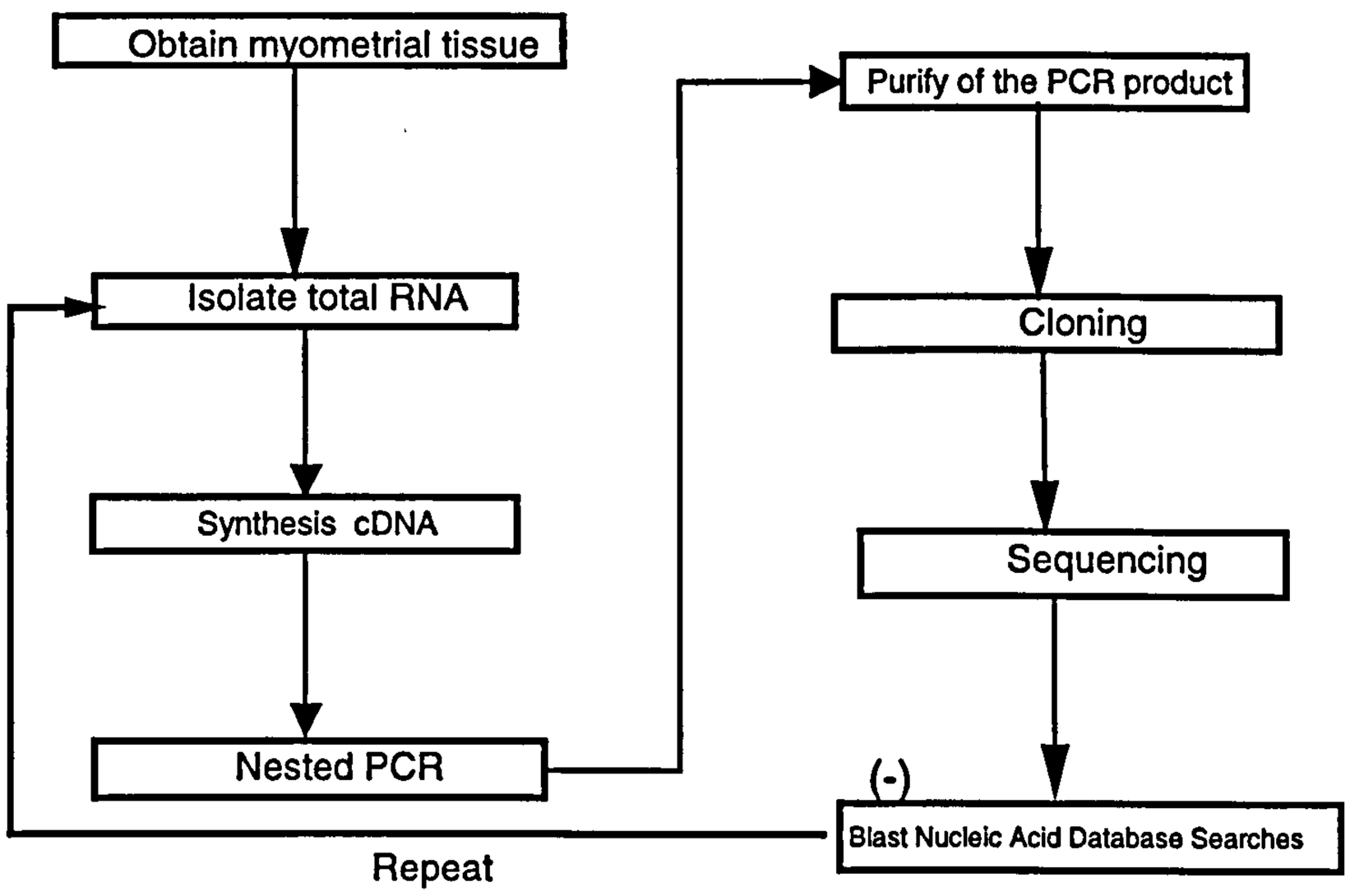


Figure 3.2. Flow chart to illustrate the procedure used to clone CRH-R subtypes from human myometrium.

3.2 Amplifying and cloning strategy

A schematic representation of the cloning strategy is shown in Figure 3.3. Polyadenylated RNA was isolated from pregnant (n=10) and non-pregnant (n=8)

myometrium tissues and myometrial cell cultures and used as a template to reverse transcribe cDNA. This reverse transcription reaction generates a cDNA template for PCR amplification. Specific primers for the CRH-R1 cDNA cloning have been described in Section 2.18. Oligonucleotide primers were based on the known sequence of the human brain CRH-R1α cDNA.

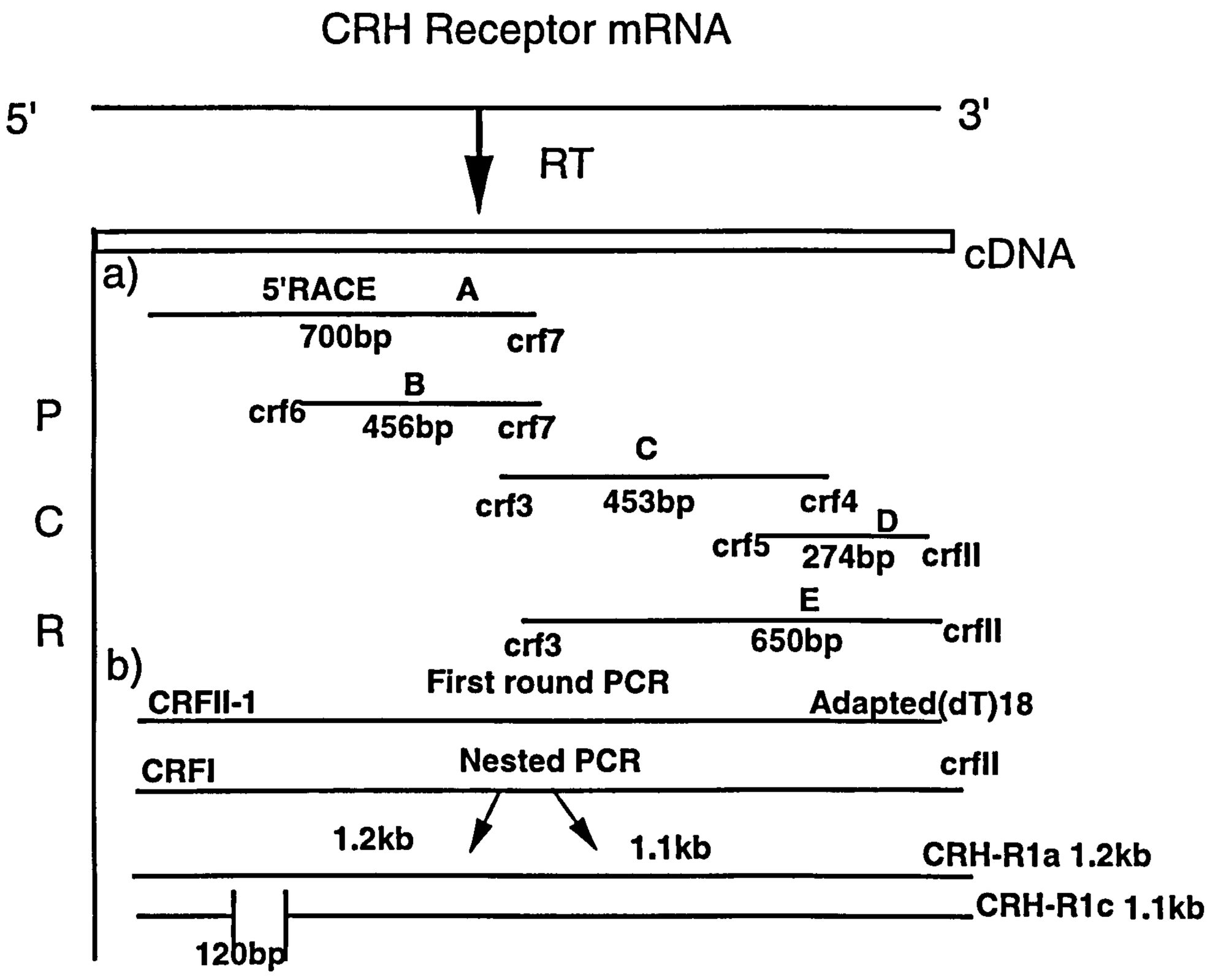


Figure 3.3. Schematic representation of the RT-PCR for amplifying strategy a) RT-PCR was amplified for different fragments of CRH-R1 using different primer pairs. b) A nested RT-PCR was used to amplify the full-length CRH-R1 subtypes.

Initially, products were amplified using different primer pairs (crf 6/crf 7, crf 3/crf 4, crf 3/crf II and crf 5/crf II). To confirm the sequence of the 5'region cDNA of CRH-

R1, 5'RACE was used to amplify the 5'-cDNA for the purpose of sequencing. Next, in order to amplify the full length CRH-R1α cDNA, a nested PCR method was employed. In the first step, a "primary fragment" was generated with the first set of primers (1S and 1A for CRH-R1α, 1c; 2S and 1A for CRH-R1β). In the second step, each cDNA fragment was cloned with a second set of primers (2S and 2A for CRH-R1α, 1c; 5'S and 5'A for 5'-end of CRH-R1β, 3'S and 2A for 3'-end of CRH-R1β). The first reaction was then used as template for the second round of amplification.

The products of the full-length PCR reaction were cloned in two different ways. The first involved the use of the *Eco*R I and *Hind* III restriction sites incorporated into CRH-R1α 2A and 2S. The PCR products were isolated by gel electrophoresis, restriction digested with *Eco*R I and *Hind* III, then cloned into pBluescript II SK. The second method utilised the A base added on to the 5' and 3' ends of the products by *Taq* polymerase and involved the cloning of the uncut, gel extracted PCR products into pGEM-T, a T-tailed vector.

The products of the second PCR reaction were purified from a 1.0% agarose gel using "QIAquick Gel Extraction kit (Qiagen)'' ligated using T4 DNA ligase kit into plasmid pBluescript II SK or pGEM-T. The plasmid with the insert was transformed into *E.coli* strain XL-1 blue cells. Clones with the insert were identified using blue-white selection. However, this did not always function efficiently as read-through often occurred due to the short length of the product. The clones were therefore checked for the insert by restriction endonuclease digestion using *Hind*III and *Eco*RI. The positively selected clones were subcultured. Plasmid DNA was then extracted using "QIAprep spin miniprep kit (Qiagen)" and digested using *Hind*III and *Eco*RI enzymes. Two bands were visualised on an ethidium bromide stained agarose gel corresponding to the plasmid and the insert. After the above procedure, isolated clones were sequenced using Sequenase version 2.0 T7 DNA polymerase kit (Amersham) and with an ABI 373 DNA sequencer (Applied Biosystems). The sequence data was analysed using Blast Nucleic Acid Database Searches from the National Centre for Biotechnology Information (NCBI).

3.4 Identification of CRH-R1 subtypes in human myometrium

In order to detect the CRH-R1 subtypes, different specific primers were employed. Initially, RT-PCR was used to amplify internal cDNA fragments B (crf6/crf 7 primer pairs), C (crf 3/crf 4 primer pairs), D (crf 5/crf II primer pairs) and E (crf 3/crf II primer pairs) comprising 274, 453, 456 and 650 bp respectively. The nucleic acid sequences were shown to overlap each other and corresponded to human CRH-R1α, which are illustrated in Figure 3.4.

To rapidly identify a 5' region sequence for this receptor, two approaches were taken. The first approach was used to amplify the 5' region cDNA of CRH-R1α by RT-PCR, but it was unsuccessful. The second was to use an anchored PCR to amplify the sequence of the 5'region of the cDNA. 5'RACE was used to amplify the 5'-cDNA for the cloning. Amplification products were size-selected, subcloned, and sequenced. The nucleic acid sequence of the cDNA fragment revealed that the 700 bp fragment corresponded to the human 5' region cDNA of CRH-R1α (see Figure 3.4).

Next, in order to amplify the full length CRH-R1α cDNA, a nested RT-PCR method was employed. When primers 2S (containing the initiation codon) and 2A (containing the stop codon) were used for the nested PCR, two DNA fragments with sizes of 1.07 and 1.18 Kb were amplified from the pregnant myometrium whilst only one fragment of 1.18 Kb was amplified from the non-pregnant myometrim (Figure3.5). These DNA fragments were subcloned for DNA sequencing. The nucleic acid sequence of the fragments revealed that, in both tissues, the 1.18 Kb fragment corresponded to the human CRH-R1α whereas the 1.07 Kb fragment amplified from pregnant myometrium was from the CRH-R1c. CRH-R1 cDNA was also amplified from human pregnant myometrial cells using the same primers and PCR conditions. The CRH-R1α subtype was identified and shown in Figure 3.6.

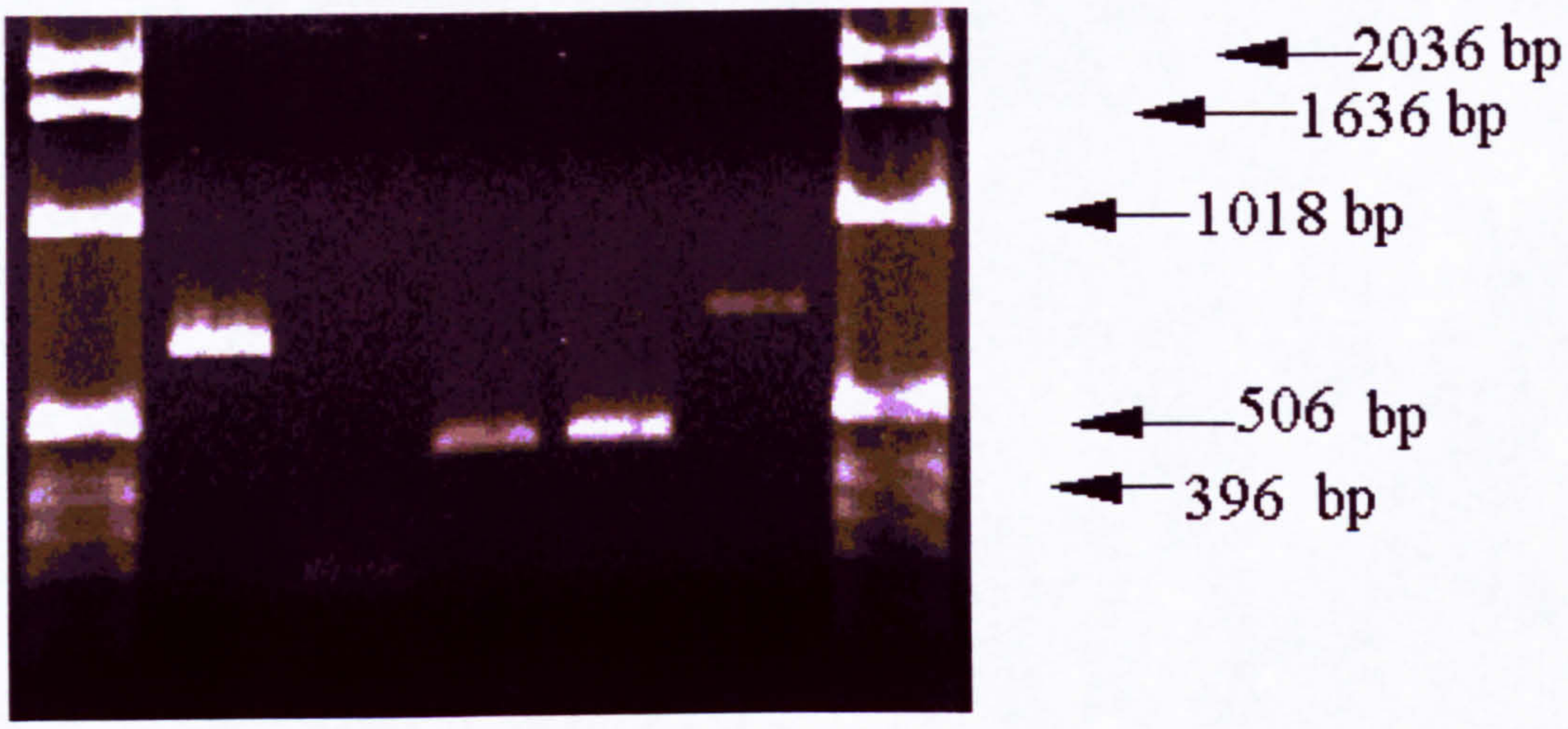


Figure 3.4. 5'RACE and RT-PCR amplification of the different fragments human CRH-R1α subtype extracted from pregnant human myometrium.
Lane 1 and 7 are DNA size makers. RT-PCR was used to amplify internal four fragments. Four cDNA fragments were generated, which have the following sizes: 650 (Lane 2), 274 (Lane 3), 453 (Lane 4) and 456 (Lane 5) bp. Lane 6 shows 5'RACE's amplification of a 700 bp fragment, which corresponds to the 5'-end CRH-R1α mRNA. The nucleic acid sequences were shown to overlap each other and corresponded to human CRH-R1α.

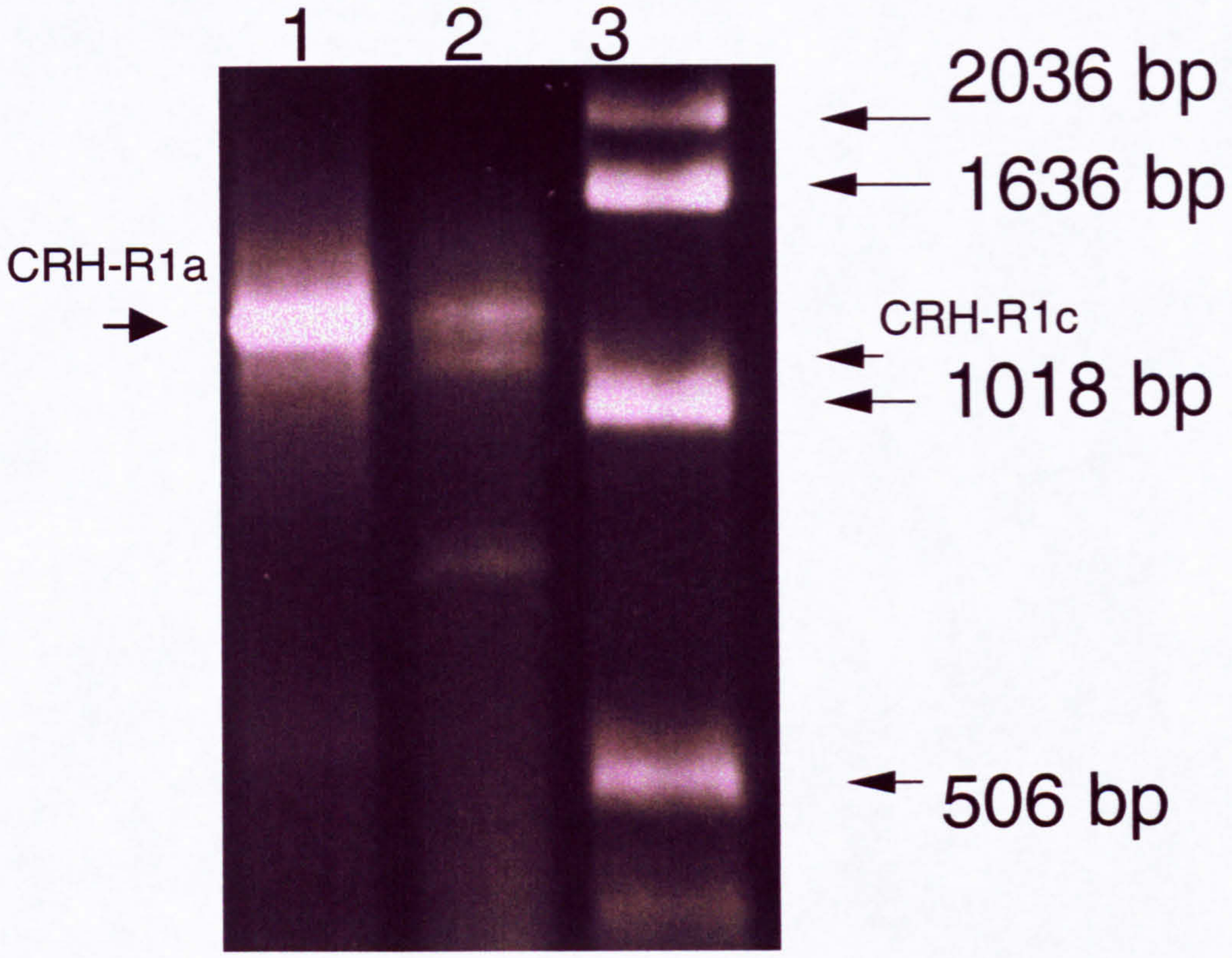


Figure 3.5 Nested PCR amplification of the human CRH-R1α and CRH-R1c from mRNA extracted from non-pregnant (Lane 1) and pregnant (Lane 2) human myometrium. The DNA sizes standards were the 1kb ladder (Gibco-BRL; Lane 3).

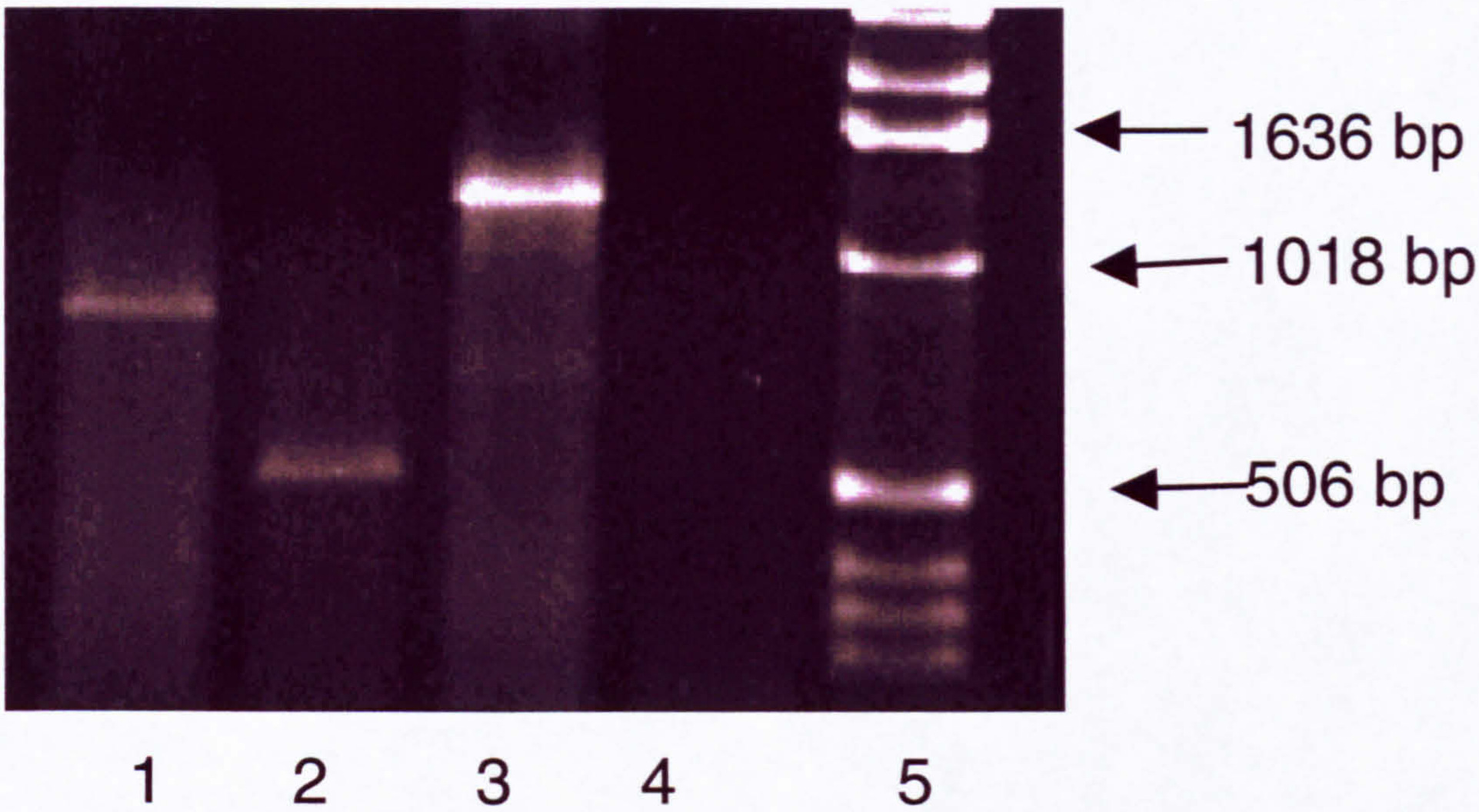


Figure 3.6 Nested PCR amplification of the human CRH-R1α and CRH-R1β from mRNA extracted from human pregnant myometrial cells (Primary culture).

For the 3'-end of the CRH-R1β, Lane 1 shows amplification of an 850 bp fragment. For the 5'-end of the CRH-R1β, Lane 2 shows amplification of a 520 bp fragment. For the CRH-R1α, Lane 3 shows amplification of a 1.18 kb fragment. Lane 4 is a negative control, distilled water was used in place of the cDNA. The DNA sizes standards were the 1kb ladder (Gibco-BRL; Lane 5).

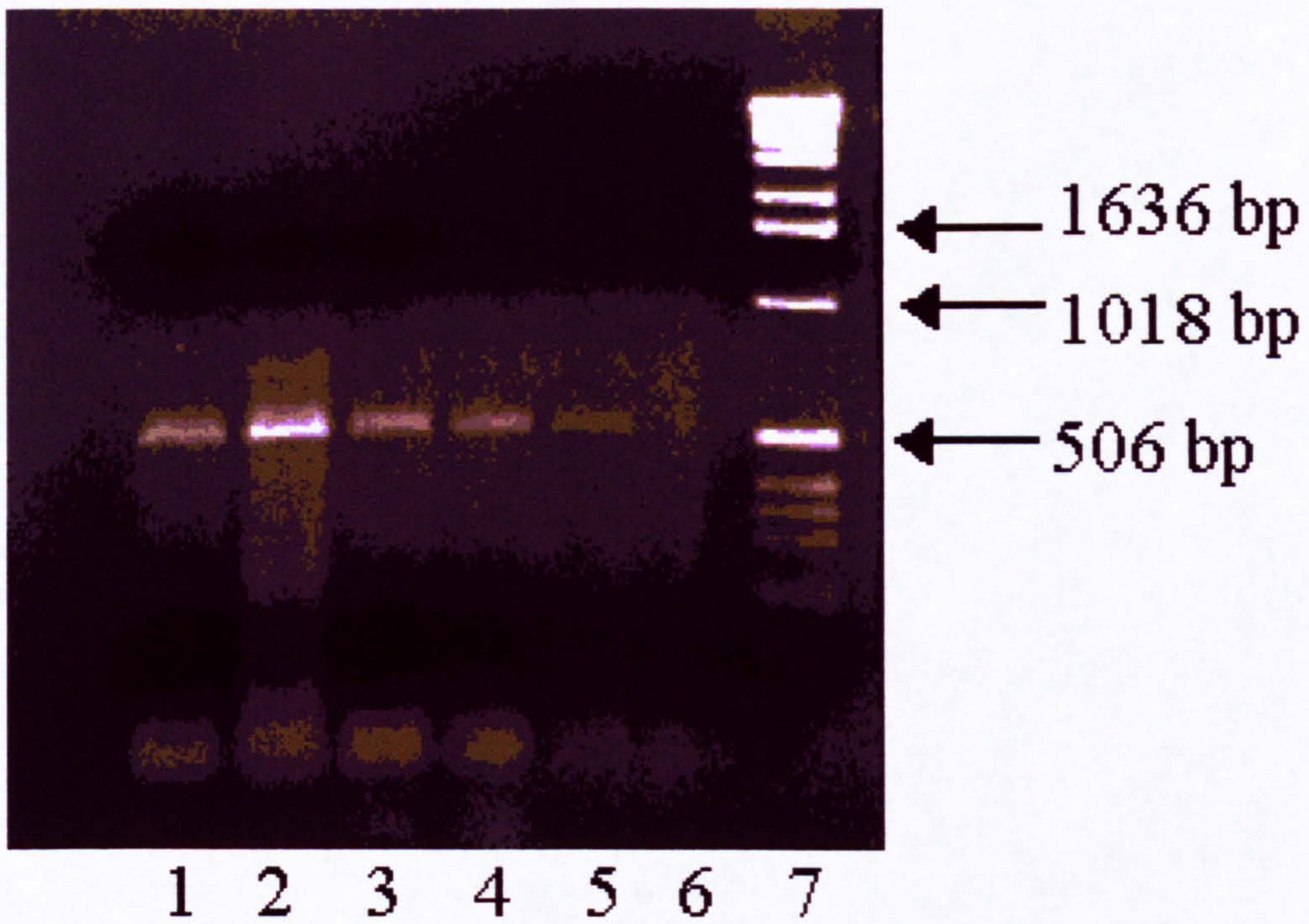


Figure 3.7 (A) Nested PCR amplification of the human CRH-R1β (5'-end) subtype.

Specific primers for the 5'-end region the mRNA were used. Human myometrium from non-pregnant (Lane 4 and 5) and pregnant (Lane 1, 2 and 3) uterus were used. For the 5'-end of the mRNA, Lane 1, 2, 3, 4, and 5 show amplification of a 520 bp fragment when compared with the DNA size marker (Lane 7). Lane 6 is a negative control, distilled water was used in place of the cDNA.

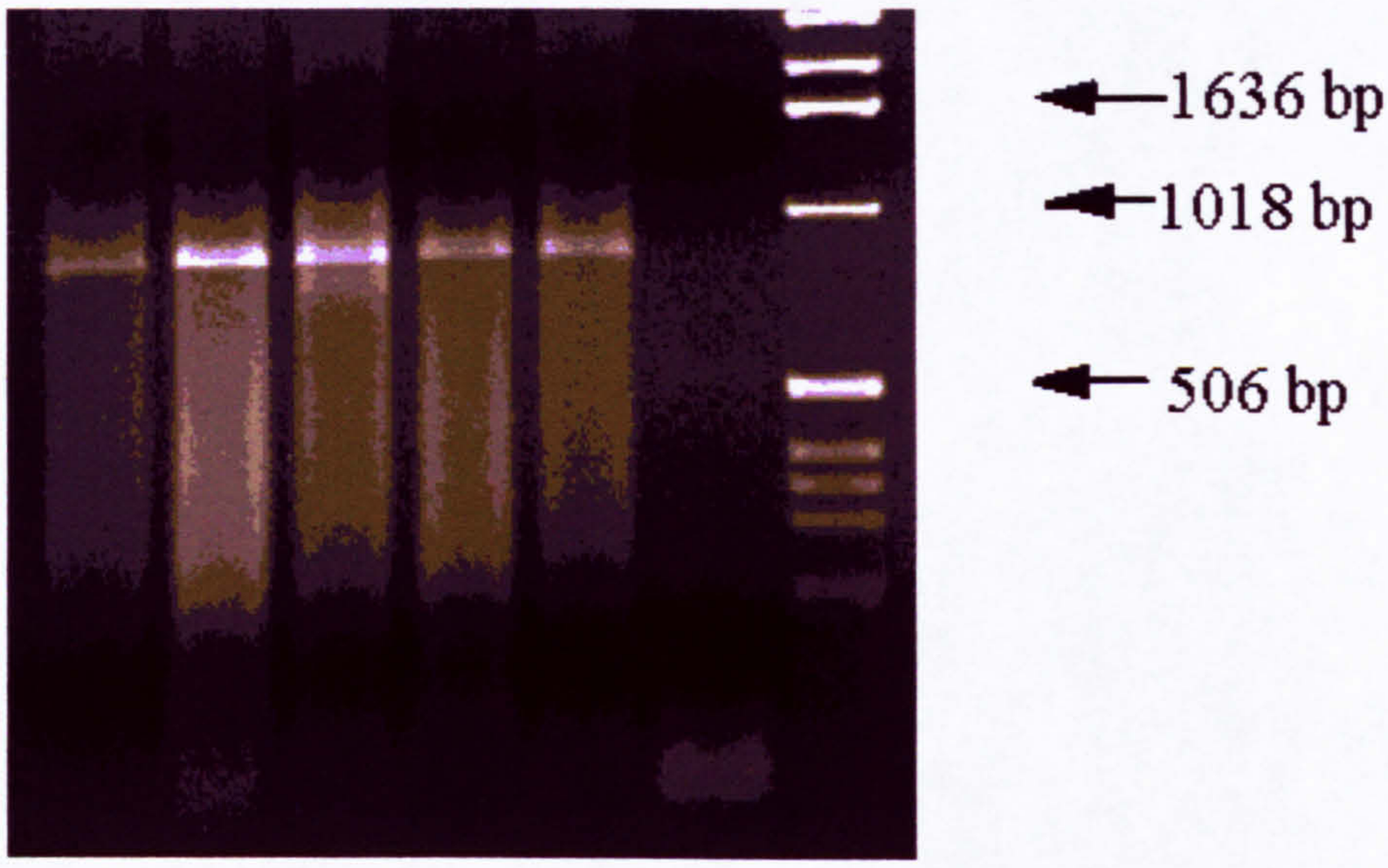


Figure 3.7 (B) Nested PCR amplification of the human CRH-R1β (3'end) subtype.

Specific primers for the 3'-end region the mRNA were used. Human myometrium from non-pregnant (Lane 1 and 2) and pregnant (Lane 3, 4 and 5) uterus were used. For the 3'-end of the mRNA, Lane 1, 2, 3, 4, and 5 show amplification of an 850 bp fragment when compared with the DNA size marker (Lane 7). Lane 6 is a negative control, distilled water was used in place of the cDNA.

In contrast, when the CRH-R1β specific primers, designed from the region of the mRNA encoding the 29 amino acid insert of the first intracellular loop of CRH-R1β, were used in a nested PCR (with primers 5'S and 5'A for the 5'-end, and 3'S and 2A primers for the 3'-end). DNA fragments of 520 and 850 bp length were amplified from both pregnant and nonpregnant tissues. Sequencing confirmed that these two bands corresponded to the CRH-R1β mRNA (Figure 3.7). The same method was used to identify CRH-R1β mRNA in the human pregnant myometrial cells. Sequencing of 520 bp and 850 bp fragments of CRH-R1β showed 100% homology with the reported sequence of the CRH-R1β gene (Figure 3.6).

3.5 RNase protection assay

To confirm the results obtained from RT-PCR, the method of RNase protection assay was used. The RNase protection assay is a sensitive method for detecting specific RNA transcripts using radiolabelled riboprobes (Zinn, *et al.*, 1983). Probes were made by synthesising antisense RNA from a DNA template using SP6 (SP6 RNA polymerase transcription initiation site), or T7 (T7 RNA polymerase transcription initiation site) RNA polymerase in the presence of at least one ^{32}P-labelled ribonucleoside triphosphate, as described in Section 2.2.15.3. Hybridisation of the probe to target RNA and RNase digestion are described in Section 2.2.14. Gel electrophoresis is described in Section 2.2.2.3.

The CRH-R1 probe was synthesised from a 450 bp 5' fragment of CRH-R1 cDNA in pBluescript SK, linearized with *Bam*H I. The CRH-R2 cRNA riboprobe was produced from a 410 bp cDNA fragment of the CRH-R2 subcloned into pBluescript SK and linearized with *Xho* I. Both CRH-R1 and CRH-R2 riboprobes were that directed against the 5' region of their respective receptors, covering the sequence up to the third presumed transmembrane region (Figure 3.8).

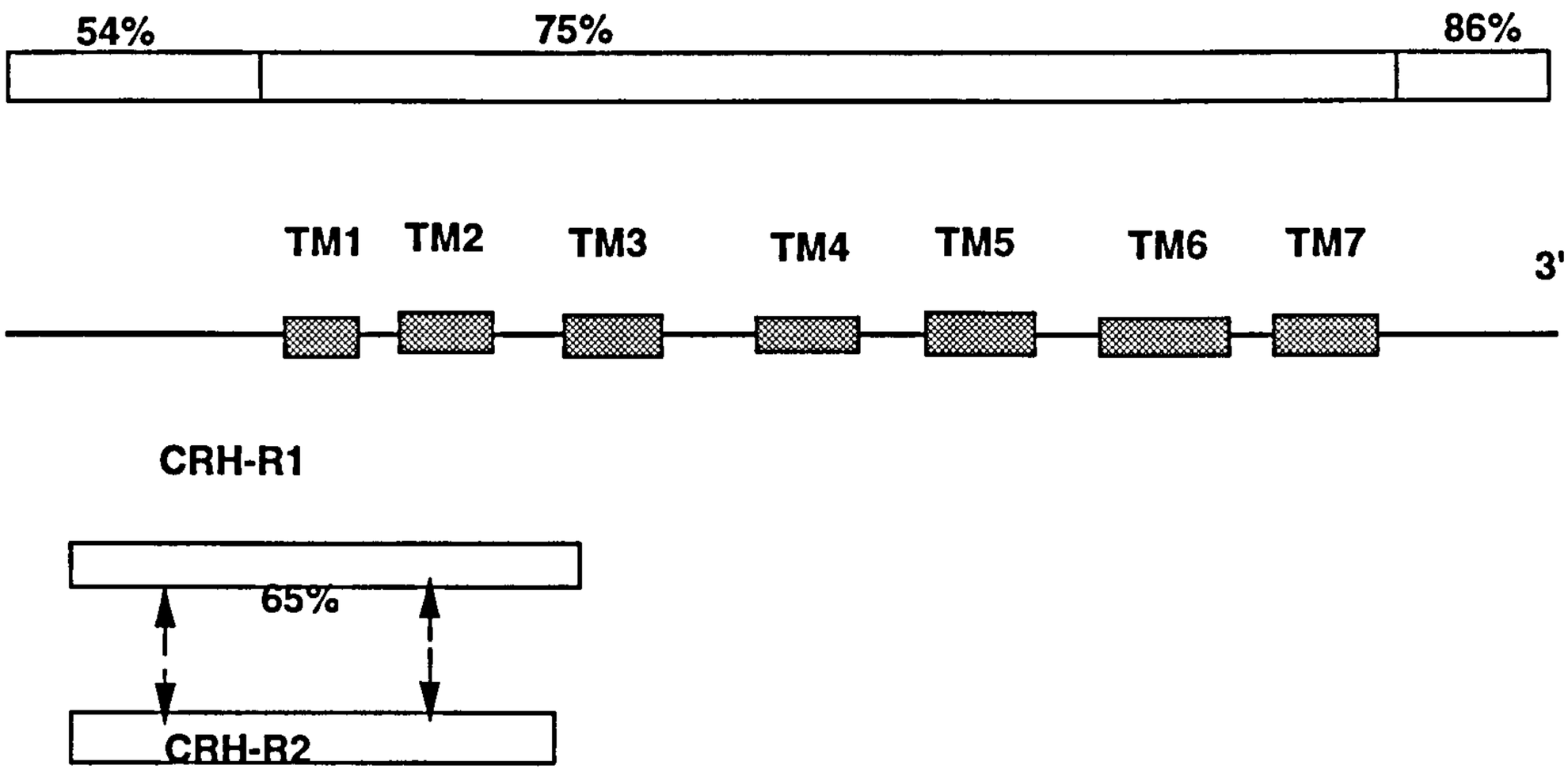


Figure 3.8 Schematic illustration of nucleotide sequence homology between CRH-R1 and CRH-R2 across the coding region of the receptors.

Lower bars indicate the region of the receptors against which CRH-R1 and CRH-R2 cDNA probes were designed.

The approximate nucleotide homology between the two probes is 65% in this region. In preliminary experiments cRNA probes directed against the 3' region of the CRH-R2 apparently labelled both CRH-R1 and CRH-R2 mRNAs whereas the two mRNA species could be clearly separated by 5' specific probes under similar hybridisation conditions.

The RNase protection assay was employed to study CRH-R mRNA expression in the human myometrium. Labelled probes were hybridised with 10-30μg of total RNA isolated from human pregnant myometrium. But it was unsuccessful. The CRH-R is present in human myometrium, but blew the level of detection by the RNase protection assay technique.

3.6. Discussion

CRH-R1 subtypes were identified in human myometrium by using RT-PCR. Initially, RT-PCR was used to amplify for different fragments of CRH-R1. To rapidly identify a 5' region sequence for this receptor two approaches were taken. The first approach was used to amplify the 5' region cDNA of CRH-R1α by RT-PCR, but it was unsuccessful. The second was to use an anchored PCR to amplify the sequence of the 5'region of the cDNA. 5'RACE is a technique that facilitates the isolation and characterisation of 5'region from low-copy messages. The second method has been reviewed by both Frohman (1988) and Loh (1989) and typically requires the design of at least three degenerate oligonucleotides. The first two oligonucleotides are used to create an enriched "template" for the second round of PCR, which then requires additional oligonucleotides corresponding to target DNA created in the first round. Each new oligonucleotide requires the presence of a corresponding sequence in the target gene and, therefore, increases the risk of missing related genes. To reduce this risk, 5'RACE PCR was combined with the nested degenerate oligonucleotides. This procedure reduces the number of necessary degenerate oligonucleotides to two, thus increasing the probability of identifying sequences related to the initial template design. Initially, a single degenerate gene-specific antisense oligonucleotide in combination with a poly(dG) oligonucleotides was used to selectively amplify the first-strand human myometrium dC-tailed cDNA. Next, nested gene-specific antisense oligonucleotides were used in combination with the original degenerate oligonucleotide to amplify DNA fragments from the first reaction. Next, a nested PCR method was employed.

This study has demonstrated that human myometrium and cultured myometrial cells (Figure 3.6) express CRH-R mRNA and extended previous observations demonstrating specific CRH receptors in human myometrium (Hillhouse *et al.*, 1993). The CRH-R1α and 1β subtypes were found in both nonpregnant and pregnant human myometrium at term before the onset of labour as well as human pregnant myometrial cells. In contrast, the CRH-R1c subtype was only found in pregnant myometrium and cells. It is possible that CRH-R1c is present in non-pregnant tissue but is below the level of detection by the RT-PCR technique. CRH-R1c was originally cloned and

characterised from the hippocampus and cerebellum of a normal 2-year-old female (Ross *et al.*, 1994). This is the second time that the CRH-R1c variant has been detected outside the central nervous system. Our laboratory has identified the full length of the CRH-R1c subtype in placental and fetal membranes (Karteris *et al.*, 1998). The function of CRH-R1c remains obscure. The exon deletion means that the receptor has a short N-terminal domain which does not bind CRH well. It is possible that this receptor may have a different ligand.

The presence of multiple CRH-R subtypes in the human pregnant myometrium as identified in this study suggests distinct functional roles for each receptor during pregnancy and raises the possibility of multiple roles for CRH and/or related peptides. Although we do not know which subtypes are more important, our data on the expression of CRH-R mRNA in myometrium support the concept that CRH plays a role in uterine physiology.

Alternative splicing is a widespread mechanism for producing variant forms of proteins from a single gene. Four subtypes of CRH-R1: 1α, 1β, 1c, and 1d (Grammatopoulos *et al.*, 1999), were found in the pregnant human myometrium at term before the onset of labour. Three human calcitonin receptor (CT-R) isoforms have been reported. The largest CT-R cDNA has a 16-amino-acid insertion in the first putative intracellular domain of the deduced protein (Albrandt *et al.*, 1995). This receptor with 16 amino acids inserted has been shown to have lower affinity and decreased G protein coupling efficiency to adenylate cyclase. This insertion is located between exon 5 and exon 7 of the CRH-R1 gene. The location of the 16 amino acids inserted within the first intracellular loop of the CT-R corresponds exactly to the location of the 29 amino acids inserted in the CRH-R1β. Another CT-R cDNA has a 14-amino-acid deletion in the seventh transmembrane domain of the deduced protein (Shyu *et al.*, 1996). The calcitonin receptor family is of particular interest since it appears to have a similar isoforms profile to the CRH-R1 family; both receptor have splice variants that contain inserts in the first intracellular loop and exon deletion in the seventh transmembrane domain. In addition, the existence of alternative splicing events indicates that there may be other unknown, splice variants of CRH-R1.

In rats, a CRH-R1 frame shift mutant encoding a 224 amino acid protein was identified (Chang *et al.*, 1993). This mutant lacks the entire third and parts of the fourth putative transmembrane domain. At the present, nothing is known about the functional properties of the mutant receptor protein. In contrast to rodents, this splice product has not been identified in humans.

Chapter 4 Amplifying and cloning CRH-R2α, 2β, and 2γ from human myometrium

4.1 Introduction

A second molecularly distinct CRH-R has been identified in the rat (Lovenberg *et al.*, 1995a), mouse (Perrin *et al.*, 1995, Stenzel *et al.*, 1995, Kishimoto *et al.*, 1995), and human (Liaw *et al.*, 1996). In the rat at least, there are two known forms of the second CRH-R. CRH-R2α is a 411 amino acid protein, which is expressed almost exclusively in brain. The alternatively spliced form, CRH-R2β, has the first 34 amino acids of CRH-R2α replaced with a 54 amino acid sequence, giving a full-length receptor of 431 amino acids. CRH-R2β shows the highest level of expression in the heart and skeletal muscle with lower level expressed in brain, lung, and intestine (Lovenberg *et al.*, 1995b). In the mouse, only the homologue of the rat CRH-R2β form has been identified (Perrin *et al.*, 1995, Stenzel, *et al.*, 1995, Kishimoto *et al.*, 1995).

The human CRH-R2 family has additional diversity in three isoforms, which utilise alternative amino termini. The three splice variants appear to diverge at the corresponding position of human CRH-R2 where the first intron occurs. CRH-R2 is expressed in the form of three functional splice variants, 2α (411 amino acids), 2β (438 amino acids) and 2γ (397 amino acids) (Liaw *et al.*, 1996, Valdenaire *et al.*, 1997, Kostich *et al.*, 1998). CRH-R2α and 2β are co-expressed in peripheral organs and the central nervous system (Valdenaire *et al.*, 1997), whereas CRH-R2γ, which has been isolated only from humans, was found only in the brain (Kostich *et al.*, 1998).

In order to detect the CRH-R2 subtype expression in human myometrial biopsies (pregnant and non-pregnant) and myometrial cell cultures, PCR primers were designed which were specific for the CRH-R2 subtype and used in nested RT-PCR.

4.2 Methods for amplifying and cloning CRH-R2

Sense primers were designed from the previously reported sequence. Antisense primers were designed from a C-terminal region shared by three splice variants common to human. Specific primers for the CRH-R2 cDNA cloning and sequencing are described in Section 2.18. All primer sets were shown not to amplify a product from genomic DNA. A schematic representation of the cloning strategy is shown in Figure 4.1.

In order to amplify the full length CRH-R2 cDNA, a nested RT-PCR method was employed. Polyadenylated RNA was extracted from human pregnant (n=10), non-pregnant (n=8) myometrium and myometrial cell cultures and reverse transcribed in each of 20µl reaction containing 400 ng Oligo $(dT)_{18}$, 0.5 mM each of four dNTPs, 50mM Tris-HCl pH8.3, 3mM $MgCl_2$, 10mM DTT and 200 units RNase H$^-$ using Reverse Transcriptase (Gibco-BRL) for 1 hour at 42°C. The reactions were stopped by incubation at 70°C for 15 minutes. To remove RNA complementary to the cDNA 2 units RNase H was added and the incubation continued for 20 minutes at 37°C. The cDNA was used as a template for a first round PCR. The products of the first PCR reaction served as a template for the second round of amplification reaction. The products of the second PCR reaction were purified, subcloned and sequenced. These differences in the receptor pattern between pregnant and non-pregnant myometrium were consistent in all subjects studied in each group.

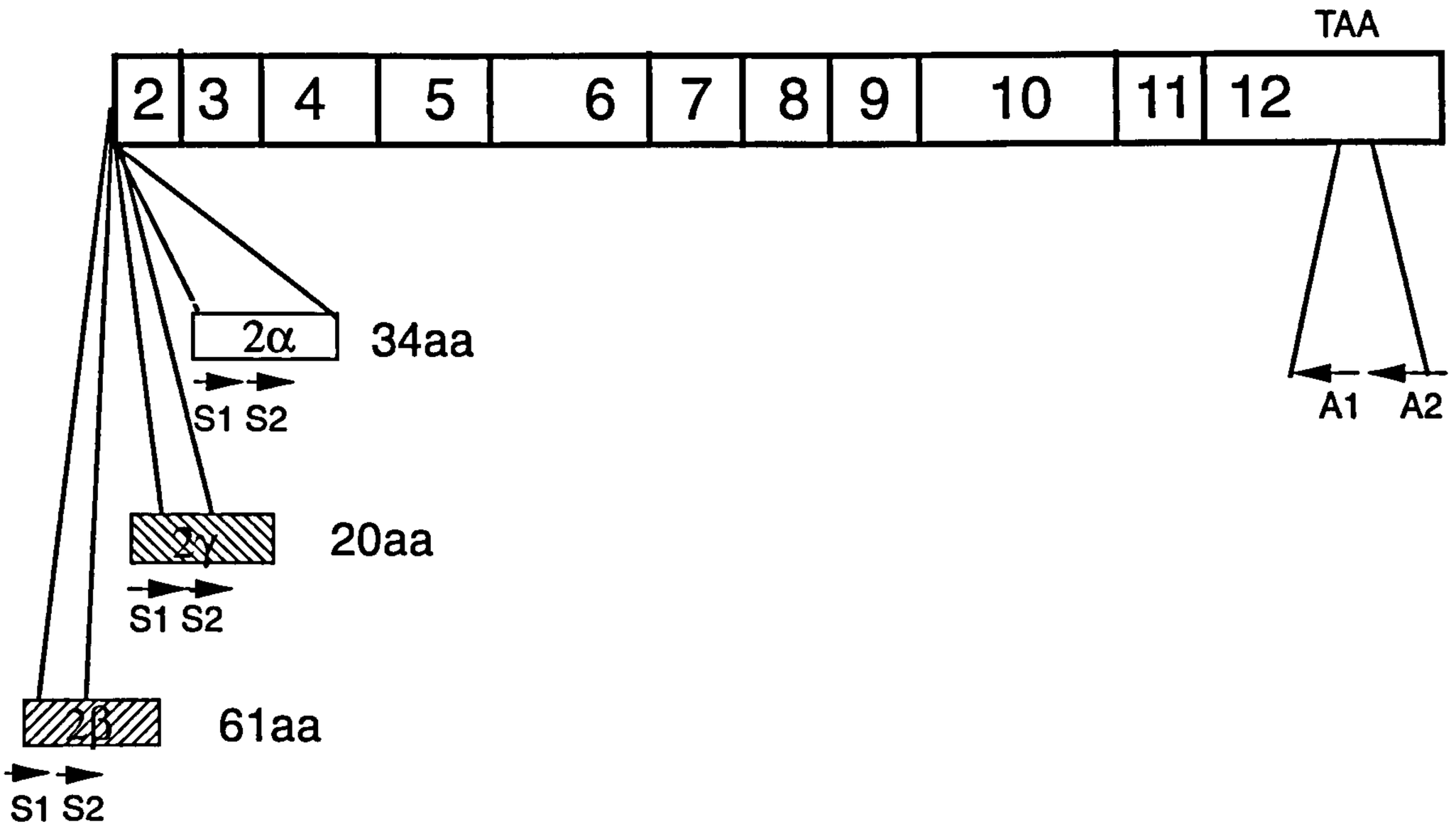


Fig 4.1 RT-PCR analysis of the human CRH-R2 mRNA isoforms 2α, 2γ, and 2β. The figure illustrates diagrammatically the approximate locations of oligonucleotides used as primers for RT-PCR analysis of the CRH-R2 mRNA isoforms. The original CRH-R2 mRNA is encoded by twelve exons, labelled 1-12. Short arrows indicate approximate locations of oligonucleotides. Note that oligonucleotides A1 and A2 in the 3'-UTR of exon 12 are common for all isoforms, whereas oligonucleotides 2α S1/S2, 2γ S1/S2, and 2β S1/S2 are specific for the different 5'-extremities. Total RNA from human myometrium was reverse transcribed using oligo(dT)18. The cDNA was then used as a template in a PCR reaction with 3'-oligonucleotide A1 and various 5'-exon-specific primers CRH-R2α, 2γ, and 2β-S1. PCR products were reamplified using a nested common 3'-oligonucleotide A2 together with nested 5'-exon -specific primers 2α, 2γ, and 2β-S2.

The products of the second PCR reaction were purified from a 1.0% agarose gel using "QIAquick Gel Extraction kit" (Qiagen) and lighted into the pGEM T-vector (Promega). *E coli* strain XL-blue 1 (200 µl) was then transformed by $CaCl_2$ with 7-10 µl of the ligation reaction. The cells were placed in 800µl L B medium and incubated at 37°C for1 hour, and plated on agar plates with ampicillin (10µg/ml), 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-gal 20mg/ml stock), and isopropythio-β-D-galactoside (IPTG

0.5M stock). The plates were incubated overnight at 37° C. After the overnight cultures, only the white colonies were picked. Plasmid DNA was then extracted from 1.5ml of the overnight culture using "QlAprep Spin Miniprep Kit" (Qiagen). The resulting plasmid DNA was then analysed using restriction enzymes *Nco* I and *Nde* I. Two bands were visualised on an ethidium bromide stained agarose gel corresponding to the plasmid and the insert. Isolated clones were sequenced using the Sequenase version 2.0 T7 DNA polymerase kit (Amersham) and with an ABI 373 DNA sequencer (Applied Biosystems). For each cDNA fragment, four independent clones were sequenced at least.

4.3 Results of RT-PCR on CRH-R2

Using CRH-R2α exon I specific primers a single strong band of 1.2 kb was amplified from human pregnant myometrium but not from the non-pregnant myometrium (Figure 4.2). Nucleotide sequencing confirmed that this single band contained the full-length CRH-R2α. Interestingly, CRH-R2α mRNA was not found in myometrial cell cultures. Because RNA was isolated from the cells on the day of confluence (on day 13± 2), we investigated the effect of the culture period on the CRH-R subtype profile. Identical CRH-R subtypes were found in the cells from the 5th day up to the 10th week of cultures (3-5th passage).

RNase protection assay was employed to detect CRH-R 2α mRNA expression in the human myometrium (as described in Section 3.3), but it was unsuccessful.

When the CRH-R2β specific primers, designed from the N-terminal region of the mRNA CRH-R2β, were used in a nested PCR, a DNA fragment of 1.3 kb length was amplified from both pregnant and non-pregnant tissues (Figure 4.3). Sequencing of this fragment showed 100% homology with the reported sequence of the CRH-R2β mRNA. Amplification of the CRH-R2β mRNA from myometrial cell cultures was accomplished employing the same RT-PCR reaction conditions and primers as those described above. Sequence of the PCR products showed full homology with the reported sequence of the CRH-R2β mRNA (Figure 4.4).

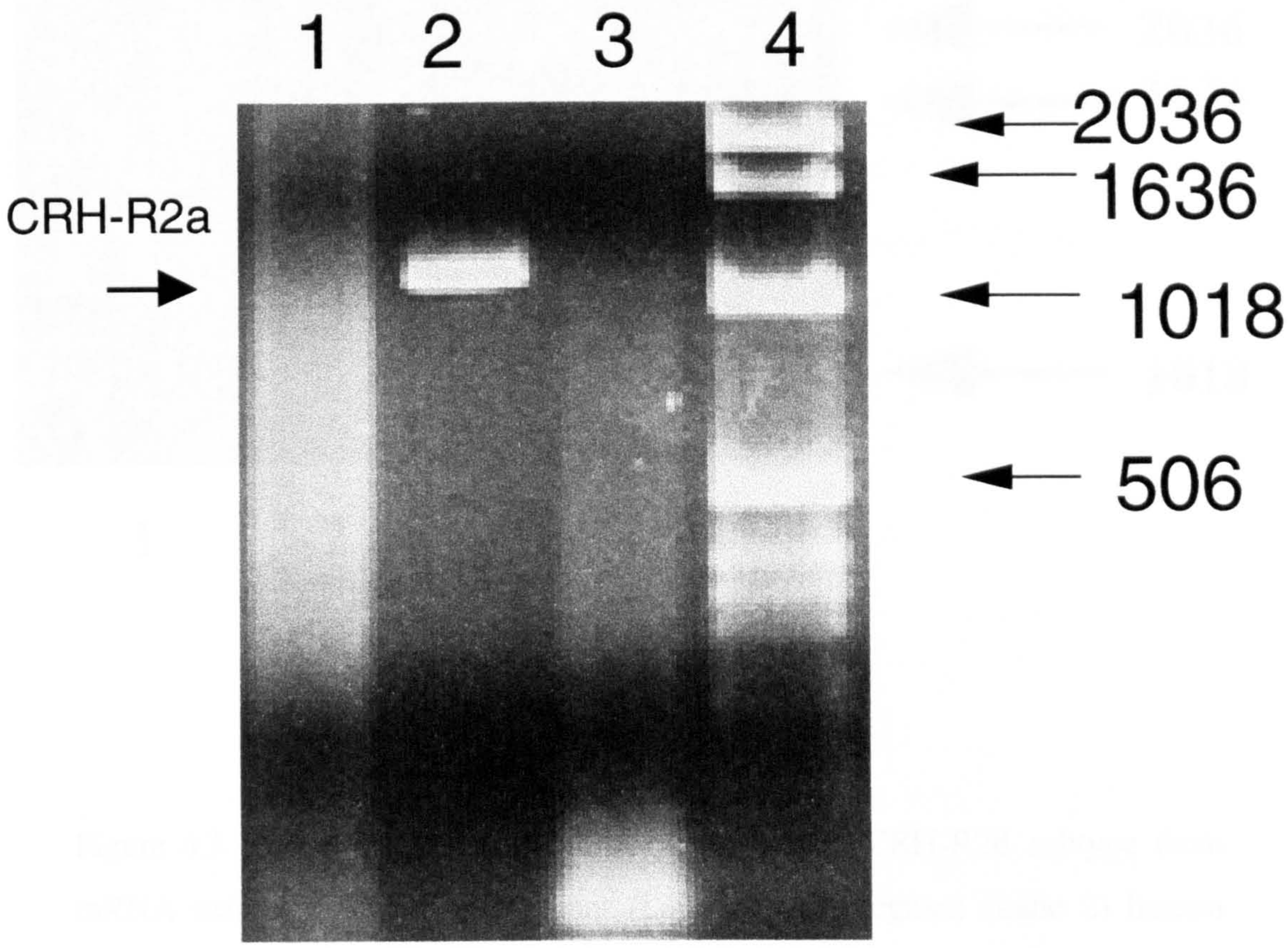


Figure 4.2 Nested PCR amplification of the human CRH-R2α subtype from mRNA extracted from non-pregnant (Lane 1) and pregnant (Lane 2) human myometrium. Lane 3 is a negative control, distilled water was used in place of the cDNA. The DNA size standards were 1kb ladder (Gibco-BRL; Lane 4).

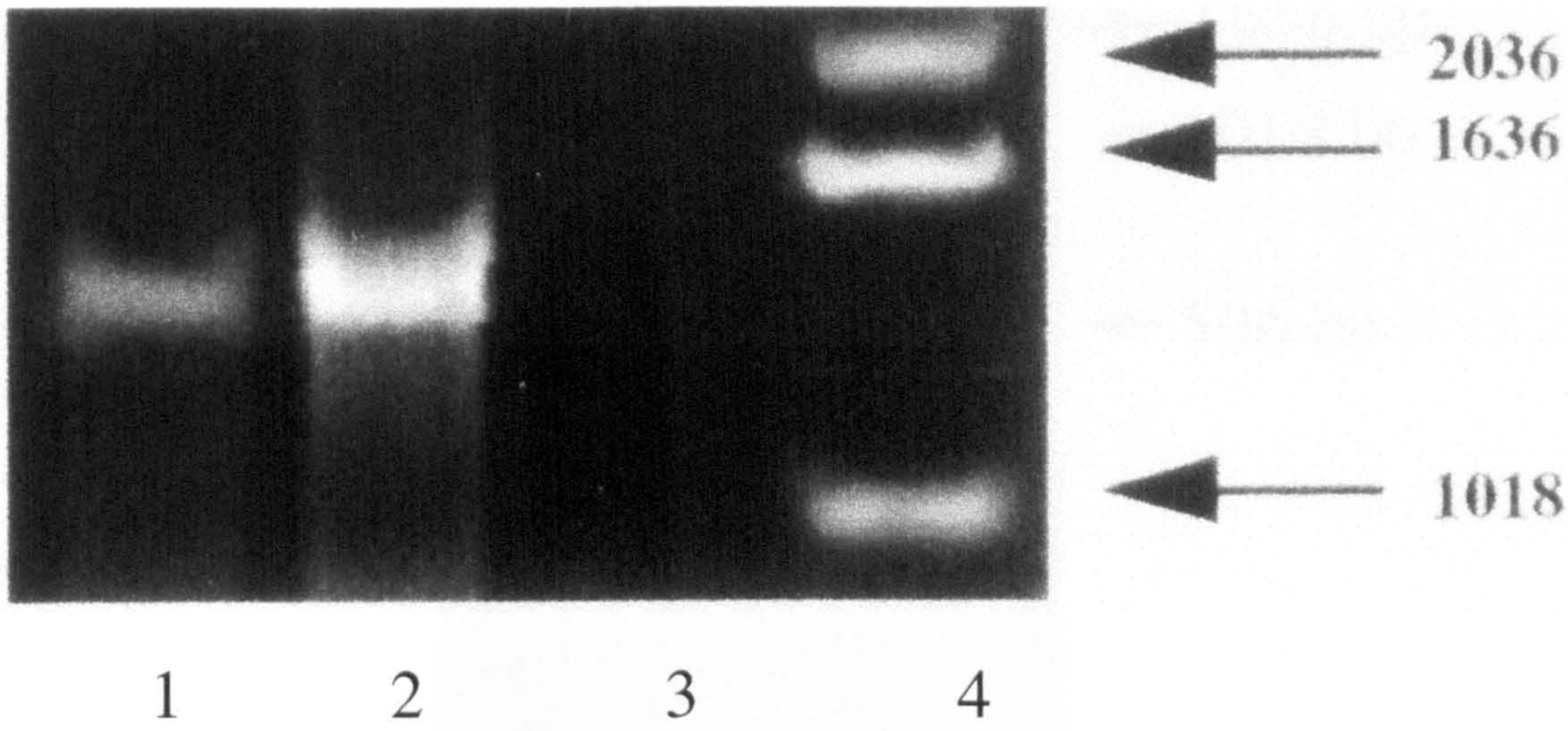


Figure 4.3 Nested PCR amplification of the human CRH-R2β subtype from mRNA extracted from non-pregnant (Lane 1) and pregnant (Lane 2) human myometrium. Lane 3 is a negative control, is a negative control, distilled water was used in place of the cDNA. The DNA size standards were 1kb ladder (Gibco-BRL; Lane 4).

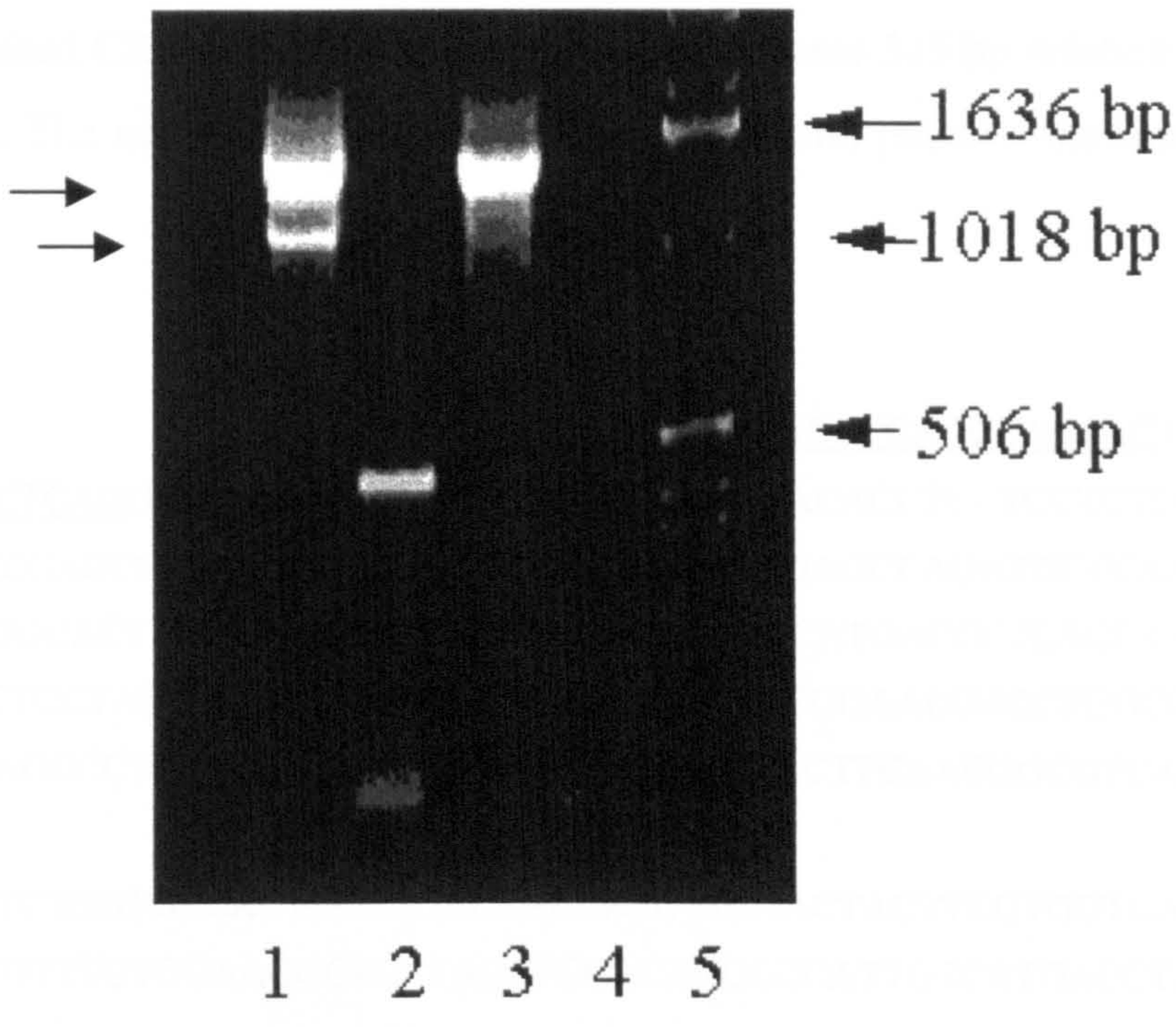


Figure 4.4 Nested PCR amplification of the human CRH-R2β (315 bp deletion) subtype from mRNA which was extracted from non-pregnant myometrium (Lane 1).

Lane 1 shows amplification of 1350 and 1035 bp fragments which correspond to the CRH-R2β and CRH-R-R2β (315 bp deletion) mRNAs. (→) Represents the CRH-R2β and CRH-R2β (315 bp deletion, mutant). The identities of the fragments were confirmed by nucleotide sequencing. Lane 2 shows first PCR product. Lane 3 shows amplification of a 1350 bp fragment of the CRH-R2β mRNA from human pregnant myometrial cells (Lane 3). Lane 4 is a negative control. Lane 5 is a DNA size marker.

In addition, when the primers were adopted, which were designed to identify the full-length CRH-R2β, we also observed a second band at 1.0 kb from non-pregnant tissue (Figure 4.4). The nucleic acid sequence of the isolated CRH-R2β is identical to the previously published CRH-R2β mRNA, except for an in-frame 315 bp deletion of exon 3-5 (Figure 4.5). The mutant lacks the entire first and second putative transmembrane domain.

A S1→

S2→ TCCAGTCCCTAACCCCAGCCACTACTGGC29

ATGAGGGGTCCCTCAGGGCCCCCAGGCCTCCTCTACGTCCCACACCTCCTCCTCTGCCT60

CTCTGCCTCCTCCCACCGCCGCTCCAATACGCAGCCGGGCAGAGCCAGATGCCCAAAG120

CAGCCCCTGTGGGCACTTCTGGAGCAGTACTGCCAGACCATCATGACCCTCACCAACC180

TCAGGTCCCTACTCCTACTGCAACACGACCTTGGACCAGATCGGAACGTGCTGGCCCC 240

AGCGCTGCCGGAGCCCTCGTGGAGAGGCCGTGCCCCGAGTACTTCAACGGCGTCAAGT300

↓

AACACGACCCGTCTGGTGCCACTGCATCACCACCATCTTCAACTACTTCGTGGTGACC 360

CTTCTTCTGGATGTTTGTGGAAGGCTGCTACCTGCACACGGCCATTGTCATGACCTACT 420

B Deleted fragment

↓

CGGAATGCCTATCGAGAATGCTTGGAGAATGGGACGTGGGCCTCAAAGATCAACTACTCA

CAGTGTGAGCCCATTTTGGATGACAAGCAGAGGAAGTATGACCTGCACTACCGCATCGCC

CTGTCGTCAACTACCTGGGCCACTGCGTATCTGTGGCAGCCCTGGTGGCCGCCTTCCTGC

TTTTCCTGGCCCTGCGGGAGCATTCGCTGTCTGCGGAATGTGATTCACTGGAACCTCATC

ACCACCTTTATCCTGCGAAATGTCATGTGGTTCCTGCTGCAGCTCGTTGACCATGAAGTG

↓

CACGAGAGCAATGAG

Figure 4.5. A shortened CRH-R2β mRNA which is identical to the normal CRH-R2β except for 315 bp deletion, the position of which is shown in (↓) as indicated. S1 is the first PCR primer. S2 is the second primer (underlined). B. Deleted fragment. The 315 bp fragment of CRH-R2β was deleted in the shortened CRH-R2β, the positions at which the fragment is deleted are indicated by (↓).

RT-PCR products

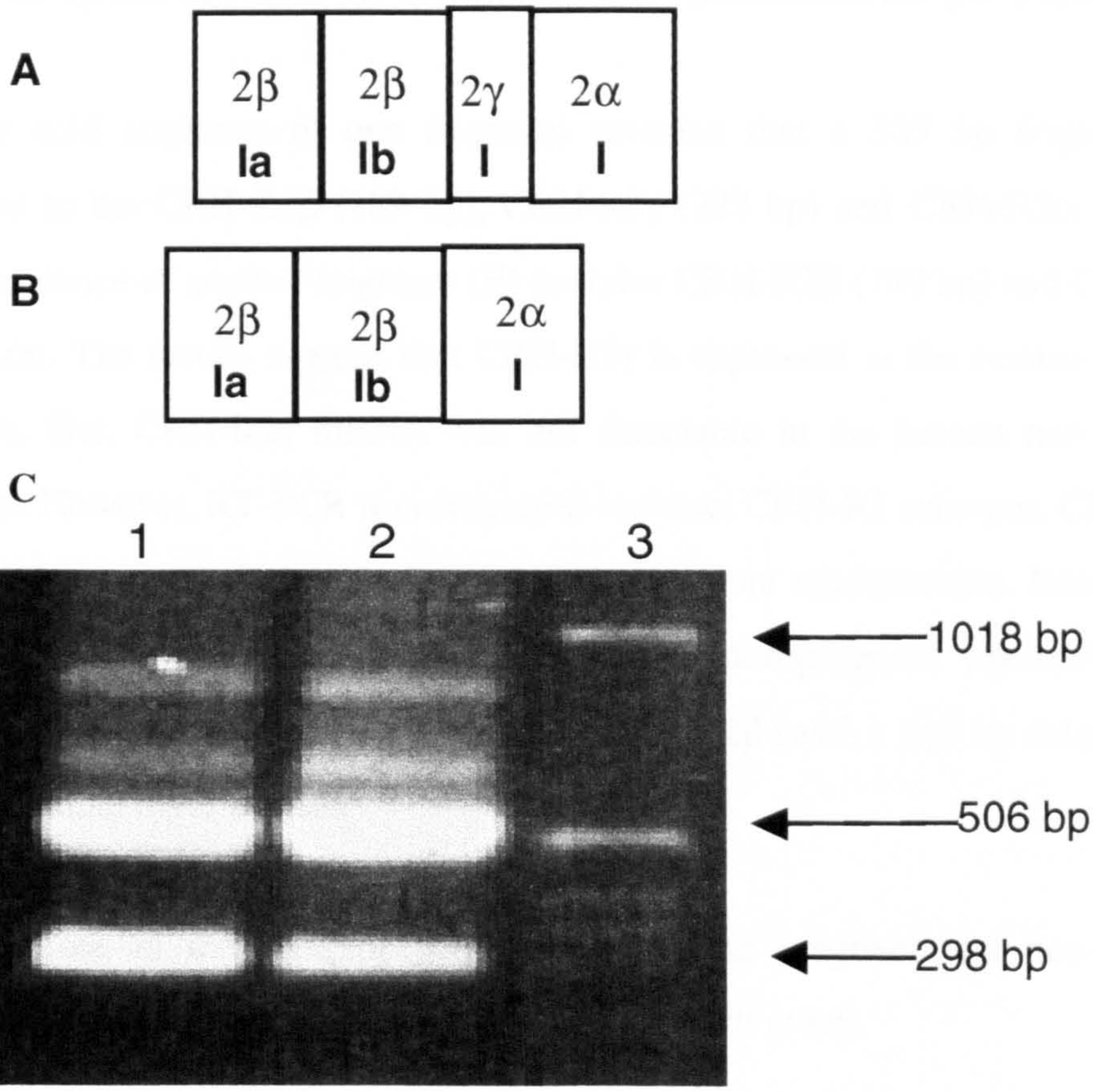


Fig 4.6 Alternative splicing of the human CRH-R2 gene and RT-PCR with CRH-R2β (exon-Ia gene) and CRH-R2α (exon-I gene) primers. The structure of the 5'flanking of the gene is shown on the top with locations of PCR primers indicated by arrows. CRH-R2β Ia 90 bp + Ib 93 bp = 183 bp (61 aa), CRH-R2γ I = 60 bp (20 aa) and CRH-R2α 112 bp (34 aa). A PCR was performed using a primer from CRH-R2β exon Ia and a primer from CRH-R2α exon I, resulting in the two bands A, B, after separation on an agarose gel (C). Lane 1 and 2 are different human myometrium sample. Lane 3 are DNA size markers. Bonds were sequenced completely, and their structures are shown in the drawing on the over boxes representing exons.

In order to detect the expression of CRH-R2γ in the myometrium, primers were designed from the reported sequence and used for RT-PCR amplification from both human pregnant and non-pregnant tissues. CRH-2γ mRNA was not detectable in the myometrium. Interestingly, a RT-PCR amplification using primers CRH-R2β (exon Ia gene) and CRH-R2α (exon I gene) resulted in two DNA fragments (A and B) of apparently 529 bp and 301 bp in the human pregnant myometrium sample (Figure 4.6).

The nucleic acid sequence of one fragment revealed that a 529 bp fragment (A) corresponded to the CRH-R2β (189 bp), CRH-R2γ (227 bp) and CRH-R2α (113 bp) exon. The sequence of another fragment (B) contains CRH-R2β (189 bp) and CRH-R2α (113 bp) exon. The results suggest that CRH-R2γ is expressed in the human pregnant myometrium. But, CRH-R2γ mRNA was not detectable in the human non-pregnant myometrium. However, RT-PCR was employed to detect CRH-R2 subtypes. CRH-R2α, 2β, and 2γ mRNA were detectable in the human pregnant myometrium. Interestingly, only one CRH-R2β was detectable in the human non-pregnant myomtrium and myometrial cell cultures. In addition, the mutant CRH-R2β (with a 315 bp deletion) was identified in human myometrium.

These differences in the receptor pattern between the pregnant and non-pregnant myometrium were consistent in all subjects studied in each group.

4.4 Discussion

We have demonstrated that CRH-R2α, 2β, and 2γ are expressed in myometrium from pregnant woman, whilst only one subtype, CRH-R2β was found in the nonpregnant myometrium and myometrial cell cultures.

Of particular interest is the finding that the CRH-R2α subtype is present in human pregnant myometrium. To date the CRH-R2α was thought to be confined solely to the CNS and there has been no report of the existence of the CRH-R2α receptor in any another human tissue. This is the first report where the CRH-R2α receptor has been

identified in a tissue other than the brain. Since there are indications that another, as yet unidentified, peptide (not CRH) is the native ligand for this CRH receptor subtype; this raises the possibility that other signals may mediate CRH-R2α effects on myometrial function. Similarly this is the first report of the expression of CRH-R2γ mRNA outside the central nervous system in humans.

The myometrial cell cultures do not express mRNA for CRH-R2α or CRH-R2γ, although the tissues of origin (pregnant myometrial tissue obtained from individuals at term undergoing elective caesarean section) were found to be positive for these CRH-R subtypes mRNA. This finding suggests that maternal factors that are components of the pregnant environment may play a major role in the regulation of CRH-R subtype expression.

We have isolated a novel CRH-R2β spliced variant mRNA from human myometrium by employing the RT-PCR technique. This variant of the CRH-R2β contains a 315 bp deletion in the exon 3-5. This mutant lacks the entire first and second putative transmembrane domain. This variant of the CRH-R2β took place just at the splicing site, suggesting that it is due to an aberrant splicing event. To date nothing is known about the functional properties of the mutant receptor protein. It is possible that the CRH-R2β variant may modulate the function of the CRH-R2β. Recently, Miyata *et al.* has identified a novel isoform of the CRH receptor, referred to as CRH-R2α-tr, from a rat amygdala cDNA library. The nucleotide sequence of the cloned cDNA has a structure of an alternatively spliced form of the CRH-R2α, which contains unspliced introns 6 and 7 in the message, and encodes a 236 amino acid truncated protein that comprises three unique transmembrane domains. These findings suggest that this truncated CRH-R is the major isoform of CRH-R2α mRNA transcripts in the amygdala and might mediate some functions of CRH pathways in the central nervous system (Miyata *et al.*, 1999). No rat CRH-R2α-tr messenger RNA (mRNA) was identified in the human myometrium. At present, there is no evidence for the existence of a CRH-R2α-tr homolog in humans.

In addition, Grammatopoulos *et al* (1995) have identified previously, by isoelectric focusing, at least five populations of CRH receptors in the human myometrium and we

do not rule out the existence of other subtypes of CRH receptors. These subtypes may arise due to alternate splicing or due to post-translational modifications (Grammatopoulos *et al.*, 1995).

Our results demonstrate the presence of multiple CRH-R mRNAs in the human myometrium (CRH-R1 mRNA, as discussed in Chapter 3). Six subtypes of the CRH receptor, 1α, 1β, 1c, 2α, 2β, and 2γ, were found in the pregnant human myometrium at term before the onset of labour, whereas only three subtypes, 1α, 1β, and 2β, were found in the non-pregnant myometrium. It is possible that the receptors are present in nonpregnant myometrial tissues but are below the level of detection by the RT-PCR technique.

This would suggest that CRH, acting via different receptor subtypes, is able to exert different actions on the myometrium in the pregnant state compared to the non-pregnant state. This was confirmed by an independent laboratory, which also showed that there is CRH-R regional distribution in the human myometrium and that CRH-R1 mRNA in the lower uterine segment was decreased during pregnancy and increased significantly in both preterm and term labour, but remained unchanged in the fundus (Stevens *et al.*, 1998).

This finding demonstrates, for the first time, that during pregnancy there is an alteration in the pattern of CRH-R subtype expression. This original observation coupled with previous studies (Hillhouse *et al.*, 1993. Grammatopoulos *et al.*, 1995, 1996) argue for a functional role for CRH and/or related peptides in myometrial function. During pregnancy CRH is produced by the placenta and feto-maternal tissues and secreted into the maternal circulation. The biological role of placental CRH is still uncertain, but it may modulate myometrial contractility, as it is able to activate specific CRH-Rs in the human myometrium. The CRH-Rs are capable of adopting a high affinity state during the latter stages of pregnancy (Hillhouse *et al.*, 1993) when they become functionally coupled to adenylate cyclase (Grammatopoulos *et al.*, 1994). This data suggests that during pregnancy CRH may play a "protective" role for the myometrium, by preventing uterine contractions. As term approaches the ability of CRH to activate myometrial adenylate cyclase is reduced (Grammatopoulos *et al.*, 1996). In particular, the ability of

CRH to activate Gsα at term is reduced. Grammatopoulos *et al.* suggested that in the term myometrium, CRH receptor function is modulated by oxytocin, leading to reduced biological activity, lower cAMP levels, and a subsequent shift in favour of contractility rather than relaxation (Grammatopoulos *et al.,* 1999).

Alternative splicing is a widespread mechanism for producing variant forms of proteins from a single gene. Generation of protein isoforms by alternative splicing is known to occur in numerous members of the G protein–coupled receptor family, correlating with differences in affinity potency, coupling efficiency, specificity, subcellular localisation or sensitivity to desensitisation of receptor. Ross *et al* reported a variant of the CRH-R1α with a deletion of 40 amino acids in the N-terminus (Ross *et al,* 1994). This form of the receptor (CRH-R1c) has previously only been identified in one individual. We now report that this variant is also expressed in the human myometrium and recent reports have also shown expression in placenta and fetal membranes (Karteris *et al.,* 1998). However, it is of interest because it has a low affinity for CRH. Interestingly, variants have also been described for which differences in interacellular regions lead to modified ligand binding, presumably as a result of altering the receptor tertiary structure. The CRH-R1α with additional 29 amino acids inserted in the putative first intracellular loop (CRH-R1β) displays a lower affinity for CRH. A number of other G protein–coupled receptors possess variants that display different signalling characteristics. Both type I vasoactive intestinal peptide (VIP) and type I pituitary adenylate cyclase-activating polypeptide (PACAP) receptors can positively couple to concurrent increases in intracellular cAMP and free Ca^{2+}, presumably through association with different G proteins (Sreedharan *et al.,* 1994, Spengler *et al.,* 1993). Spengler *et al.* described five splice variants of the PACAP receptor differing in the region corresponding to the third intracellular loop of this receptor (Spengler *et al.,* 1993). Four of these five splice variants exhibited the multifunctional signalling capability of PACAP receptors described in several cell types, activating both adenylyl cyclase and phosphoinositide phospholipase C. Structurally, the PACAP receptor exhibits sequence homology to the corticotrophin-releasing hormone, parathyroid and glucagon receptor family that have been designated as member of Class II in the GPCR superfamily. Amplification of this region (the third intracellular loop) from cDNA made from the myometrium has not indicated the presence of other isoforms of

the CRH-R.

The human CRH receptor is a member of a newly characterised family of G protein-coupled receptor that shares several characteristics structural features. It has been reported that different subtypes of CRH receptor have different efficiencies of coupling with G proteins (Chen *et al.*1993, Perrin *et al.*, 1995). Different mechanisms may be involved in order to increase myometrial sensitivity to different ligands dependent upon the presence of the subtypes of CRH receptor on the membrane of myometrial smooth muscle cells. Therefore, the high concentration of CRH in the maternal circulation, coupled with high expression of CRH receptor subtype and high cAMP production in the uterus, would favour myometrial quiescence during pregnancy.

Although the physiological role of CRH in myometrial function is not fully understood, the pattern of expression of the CRH receptor mRNA suggests that this receptor has a physiological role in the regulation of uterine function. Further studies are currently under way in order to identify the distribution of CRH receptor subtypes in uterine feto-maternal regions as well as to investigate the signals that regulate the pattern of expression of the CRH-R subtypes.

Chapter 5: Screening of a human genomic cosmid library for the 5' end of the CRH-R1

5.1 Introduction

Molecular biology has greatly aided our understanding of the expression and regulation of peptide hormone receptors. The isolation and characterisation of peptide hormone receptor genes have led to deductions about their primary structure as well as their biological activity. Study of the control regions of these genes has also given insight into factors that control both basal and hormonally regulated expressions.

Mammalian genes, unlike prokaryotic genes, contain both exons and introns. Exons are DNA sequences that appear as RNA sequence in the mature messenger RNA (mRNA) encoding a particular gene. Introns are DNA sequences that do not appear as mRNA in the mature mRNA. The first exon contains 5'-untranslated sequence. The last exon contains 3'-untranslated region (Mout 1982). At the borders of all introns there are consensus sequences that allow for RNA splicing. The consensus sequence for a splice donor at the 5'end of an intron is GT. The consensus splice acceptor at the 3'end of the intron is AG. These splice donor and splice acceptor sites in the primary DNA sequence of mammalian genes are important for accurate and efficient RNA splicing.

Exons can contain either coding sequence or non-coding sequence; the latter is referred to as untranslated sequence. These DNA sequences are present in the mRNA sequence and are important for translation initiation as well as mRNA stability. Control regions for mammalian genes can be located at any position within the gene. However, most control regions to date have been found in the 5'-flanking region of the gene. This region contains DNA sequences necessary for accurate and efficient transcription, as well as enhancer, repressor, cell-specific, and hormonal regulatory elements. Mammalian genes have a consensus TATA box approximately 20-30 bp upstream from the start of transcription. The TATA box is important for binding the transcription initiation complex, which is composed of a variety of different protein subunits. Also located within the 5'-flanking region is DNA sequences that bind

specific trans-acting factors important for basal, cell-specific, and hormonally regulated expression.

Corticotrophin-releasing hormone (CRH) is a principal neuro-regulator within the hypothalamo-pituitary-adrenal axis that co-ordinates endocrine, behavioural, and autonomic responses to stress. CRH exerts its effects by initially binding to a membrane-bound receptor. High affinity CRH binding sites has been characterised in the rat, mouse, and human. Cloning of cDNA from different species and tissues has revealed that all species investigated so far produce two receptor subtypes, CRH-R1 and CRH-R2. These receptors show differences both at the level of their expression between tissue types and in their pharmacological properties, (as discussed in Chapter 1.3.) The CRH-R cDNA from human, rat, and mouse indicates that the expressed protein is a member of the G protein-coupled seven-transmembrane receptor (GPCR) family and has been shown to couple to both adenylate cyclase and phospholipase C, stimulating cAMP production and phosphoinositide hydrolysis. The CRH-R exhibits significant homology to the parathyroid and glucagon receptor family that has been designated as members of Class II in the GPCR superfamily. This class exhibits little primary sequence similarity to the Class I GPCR, and members contain specific conserved amino acid motifs in the extracellular and transmembrane domains. In addition, genomic structures of the parathyroid and glucagon receptors suggest that most members of the Class II family differ from the Class I families by the presence of introns in both the N-terminal extracellular and the transmembrane/cytoplasmic domains.

The gene coding for human CRH-R1 consists of at least 13 exons and spans over 20 kilobases (Sakai *et al.,* 1998). The rat CRH-R1 gene has a similar origination (Tsai-Morris *et al.*, 1996). The CRH-R1 gene has been mapped to chromosome 17q12-q22 (Vamvakopoulos and Sioutopulou 1994). CRH-R1's three reported isoforms originate from the same gene by alternative splicing. So far, the sequence of the promoter or 5'-flanking and transcription initiation site of the human CRH-R1 and CRH-R2 gene is not yet understood. To understand the molecular mechanisms that direct the expression of the gene encoding the CRH-R and to analyse the structure and pattern of transcription of the CRH-R genes, it is important to clone and characterise the promoter region of the human CRH-R1 genes.

5.2 Screening and analysis of results

5.2.1 Library screening

A genomic library is a large number of DNA fragments derived from the genome, which are inserted into a vector which can be a plasmid, a bacteriophage, or a cosmid. Genomic DNA is first fractionated into sizes compatible with the chosen vector. In general, vectors are chosen based on the size of genomic DNA fragments that will be inserted. The plasmid-based vectors are useful for inserts that are less than 10 kb in size. Bacteriophage vectors are useful for inserts that are approximately 8-20 kb in size. Cosmid vectors will accept genomic DNA fragments of 50 kb and greater in length. Cosmid vectors are modified plasmids that contain λ cohesive ends sites and are packaged like a bacteriophage, but are propagated like plasmids in *E. coli.* An amplified human placental genomic cosmid DNA library (Clontech) was used for this project. The advantages of screening cosmid DNA libraries are (a) isolation of larger genes than those that can be found in other vector cloning systems; (b) isolation of linked genes on one DNA segment; (c) recovery of flanking sequences of the gene of interest, and (d) reduction of the number of colonies that need to be screened. Although cosmid libraries offer the advantage of very large inserts, these libraries are more difficult to screen than bacteriophage libraries because the cosmid vector contains plasmid sequences that may contaminate any radiolabeled DNA probe.

An amplified human placental genomic cosmid DNA library (Clontech) was used to isolate clones containing the 5'-flanking of the CRH-R1 gene. The library was constructed from 30-50 kilobase (kb) fragments of the partial *Bam*H I digest of placental DNA inserted into the *Bam*H I site of the pWE15 cloning vector. Plaque lift of 4×10^5 recombinants was hybridised with a 450-base pair radiolablled CRH-R1α cDNA probe (Oligolabelling kit, Pharmacia Biotech) As discussed in Chapter 2, the fragment was derived from the 5'-end (1-450bp, ATG translation initiation codon in exon I, which is denoted as nt +1) of a human CRH-R1α cDNA clone (Figure 5.1). Filters were washed twice for 15 minutes at room temperature with a 2 x SSC, 0.1% SDS, once with 0.1 x SSC, 0.1% SDS at 60°C for 60 minutes. Positive plaques were

identified by autoradiography and purified through a secondary and tertiary round of screening.

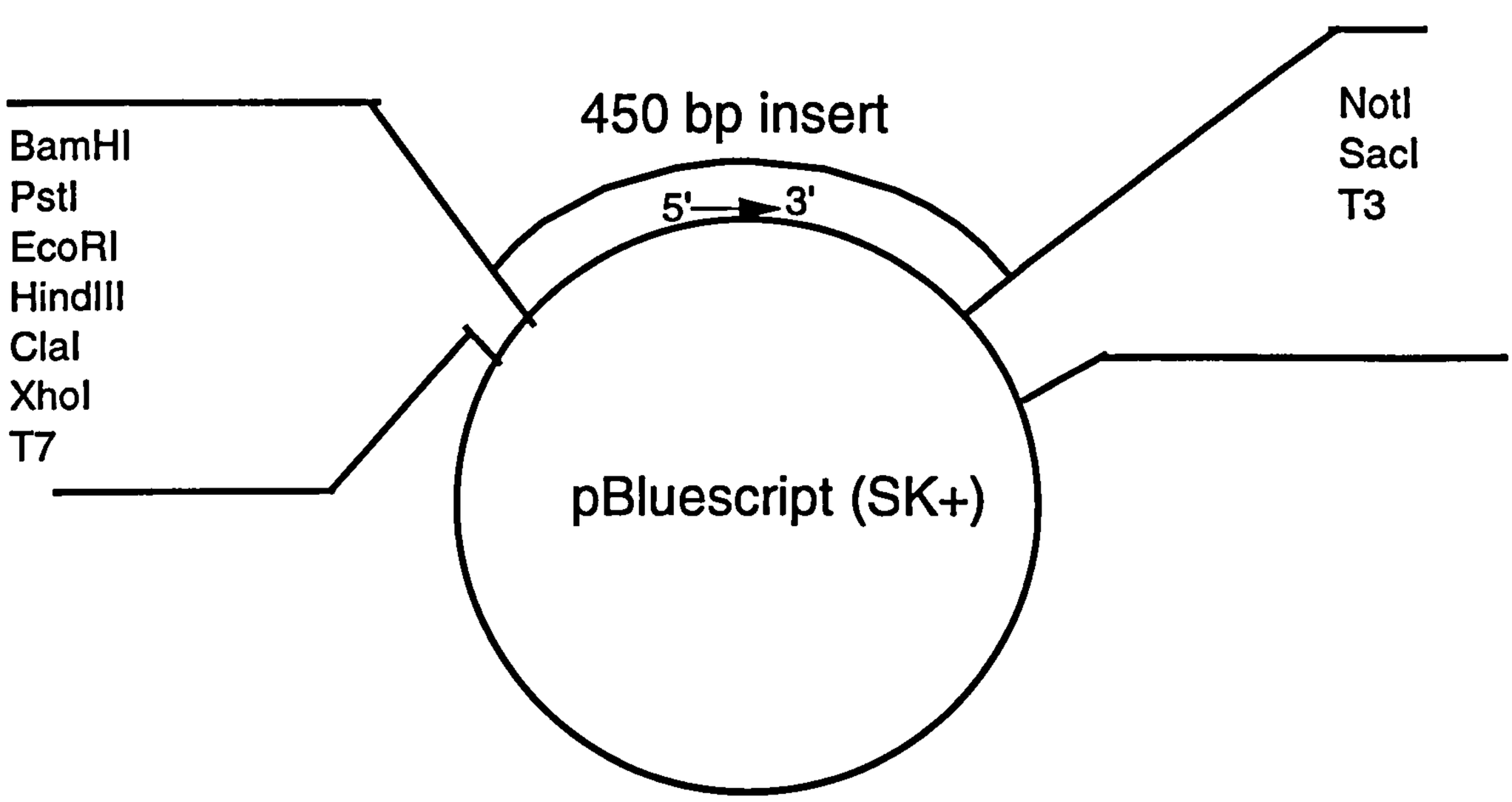


Figure 5.1 Human CRH-R1 probe inserted into the pBluescript (SK+) cloning vector. Plasmids were digested with *Xho* I and *Not* I, and purified from a 1.0% agarose gel using "QIAquick Gel Extraction kit (Qiagen).

5.2.2 Restriction digests, southern blot, and sequence analysis

Detailed restriction maps and Southern blot hybridisation should be performed with restriction enzymes known to digest cosmid vector DNA in only a limited number of locations. After a detailed restriction map is performed, the individual DNA fragments can be further subcloned and sequenced to confirm their identity. The restriction map of any genomic clone should match a southern blot of genomic DNA.

The isolated exon-containing DNA fragments were digested, subcloned into pGEM-5Zf vector (Promega) and purified by the QIAprep spin columns (Qiagen) method. The nucleotide sequence was determined by the dideoxy chain termination method with modified T 7 DNA polymerase (sequenase Version 2.0 DNA sequencing Kit Amersham) and with an automated DNA Sequencer (ABI 373A, Applied Biosystems).

From a screen of ≈ 500,000 independent clones, one positive clone was found. The clone was named J1, and was ≈ 30kb in length. Restriction digestion of the genomic clone with *Bam*H I released five fragments of approximately 4kb, 3.5kb, 3kb, 1.8kb and 900bp. The Genomic fragments were individually subcloned into a pGEM-5Zf vector. The DNA was digested with *Bam*H I and separated on a 1% agarose gel. Denatured DNA was transferred to nylon membrane (Hybond-N, Amersham) and hybridised with random prime labelled cDNA (CRH-R1α). The blots were hybridised to the labelled cDNA corresponding to exons I-II and exons I-V. Firstly, two *Bam*H I fragments 1.8kb and 4.0kb were labelled when a Southern blot was hybridised to the radiolabelled CRH-R1α cDNA probe from exons I-V (Figure 5.2). A J1-*Bam*H I 1.8kb fragment was subcloned and sequenced. Based on the sequence analysis, the fragment included exons 5-6 of the human CRH-R1α gene.

Next, A J1-*Bam*H I 4.0kb fragment was identified when a Southern blot was hybridised to the radiolabelled cDNA corresponding to exons I-II (Figure 5.3). This probe was generated from 450 bp cDNA by restriction digestion with *Pst* I and corresponded to the first 157 nucleotides of CRH-R1α.

The J1-*Bam*H I 4.0kb fragment was digested again with *Sph* I. After electrophoresis and blotting onto Hybond N, Southern blot was hybridised with a cDNA probe from exons I-II. The fragment contains a J1-*Sph* I 1.2kb signal (Figure 5.4).

A J1-*Bam*H I 1.8kb fragment was subcloned and sequenced. Based on the sequence analysis, the fragment included exons V-VI of the human CRH-R1α gene. The J1-*Sph*I 1.2 kb fragment was subcloned and sequenced. The nucleic acid sequence of the fragment revealed that the J1-*Sph* I 1.2 kb fragment corresponded to exon II of the human CRH-R1α gene.

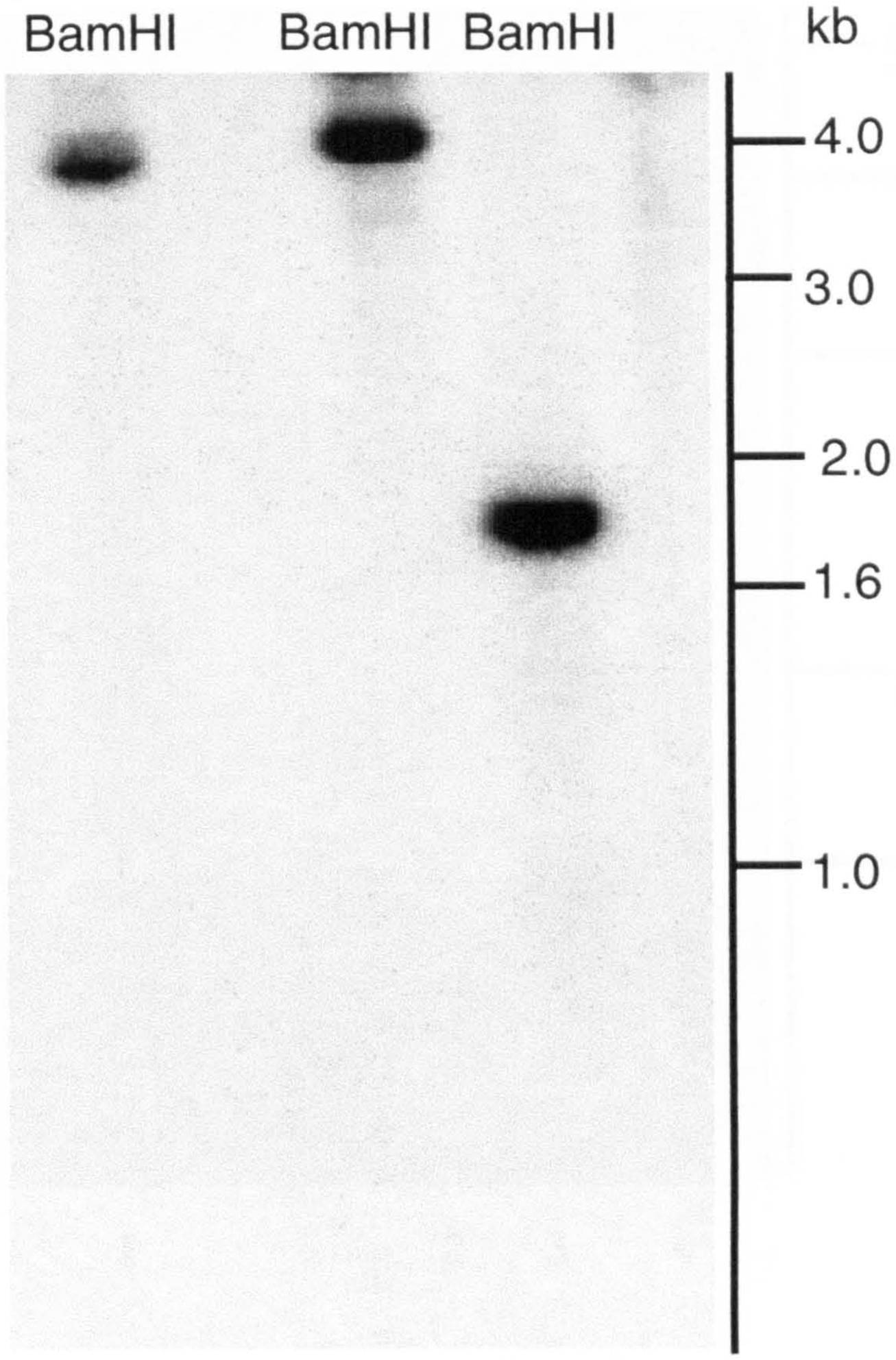


Figure 5.2 Southern blot analysis of clone contain CRH-R2α gene

The genomic clone was digested with *Bam*H I and subcloned into a pGEM-5Zf vector. The plasmid DNA was digested with *Bam*H I and separated on agarose gel. DNA was and transferred to Hybond-N. Filter was probed with the radiolabeled cDNA (CRH-R1α). This identified two *Bam*H I fragment with sizes of 1.8 and 4.0 kb, respectively.

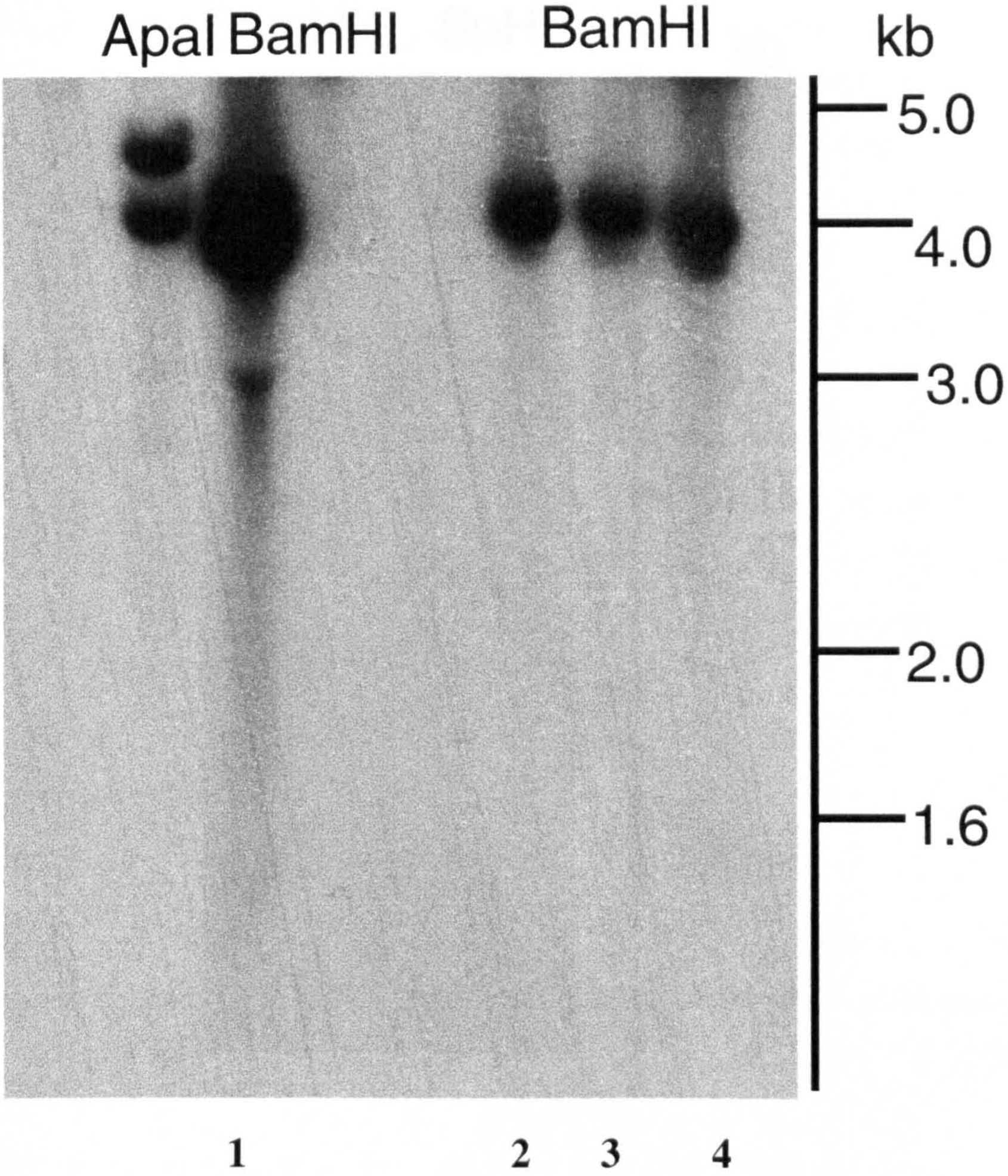


Figure 5.3 Southern blot analysis of the genomic clone fragment with the CRH-R1α cDNA probe (157 bp length)

The genomic clone was digested with *Bam* H I (lane 2, 3 and 4), separated on agarose gel, and transferred to Hybond-N. Filter was probed with the radiolabeled cDNA corresponding to the first 157 nucleotides of the human CRH-R1α. Lane 1, plasmid *Bam* H I 4kb fragment is used as a positive control. This identified a *Bam* H I 4.0 kb fragment.

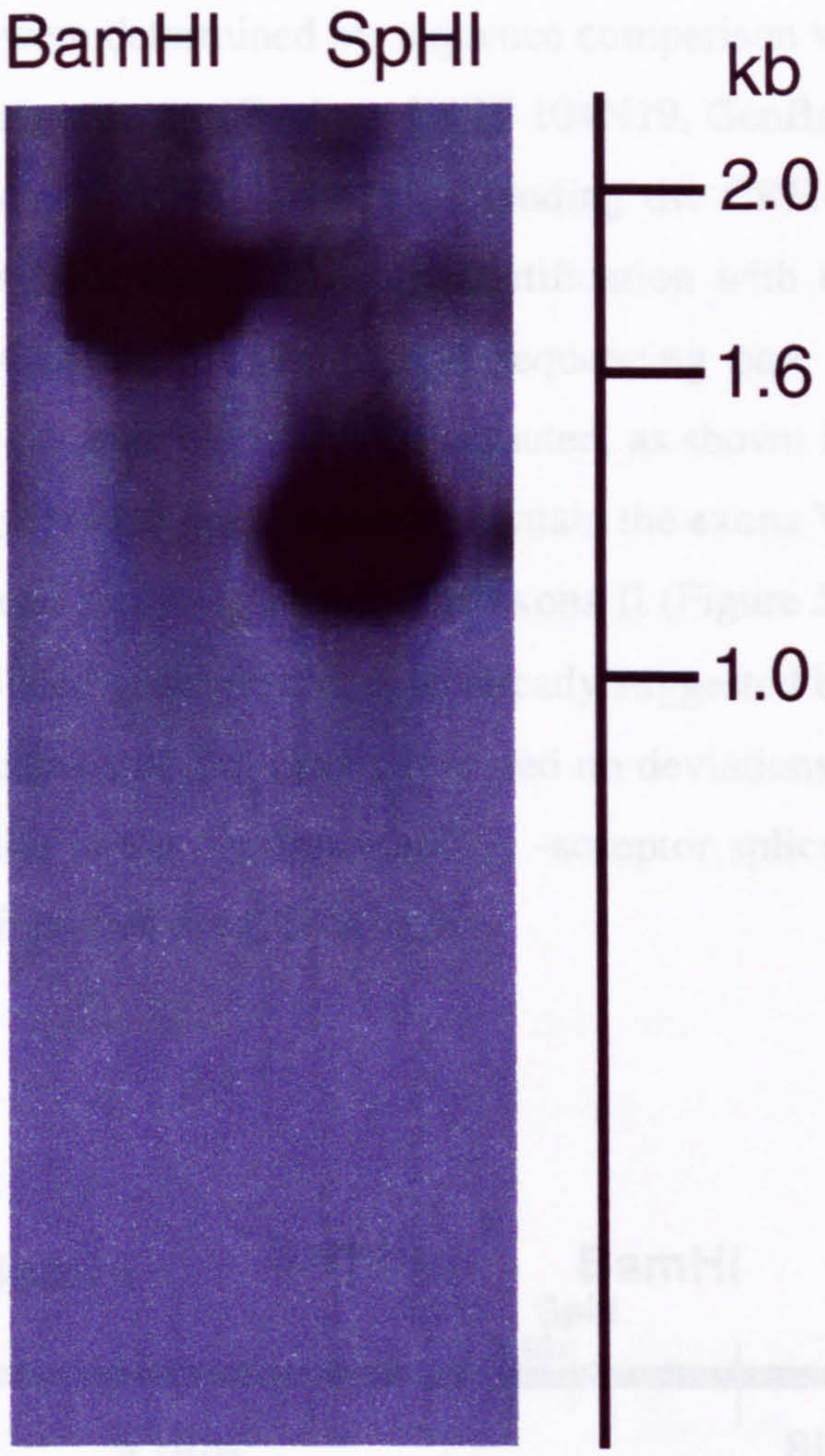


Figure 5.4 Southern blot analysis of the clone fragment containing the 5'end of the CRH-R1 gene

The J1-*Bam*H I 4.0kb fragment was digested again with *Sph* I. After electrophoresis and blotting onto a membran, Southern blot was hybridised with a cDNA probe from exons I-II of the CRH-R1 gene. The fragment contains a J1-*Sph* I 1.2kb signal (Lane 2). The plamid *Bam*H I 1.8 kb fragment was used as a positive control (Lane 1).

The nucleotide sequence of the coding region was determined on both DNA strands. Intron-exon junctions were determined by sequence comparison with CRH-R1 cDNA and Homo Sapiens Chromosome 17 (clone HCIT 104N19, GenBank No. AC003662). The DNA sequence (clone HCIT 104N19), including the CRH-R1 gene except for exon I, was made available recently. After identification with two different cDNA probes (fragments 1-450 and 1-157 bp) and sequencing part of the subclones, a restriction map of the genomic clone was constructed, as shown in Figure 5.5 A. The J1-*Bam*H I 1.8 kb fragment has been shown to contain the exons V-IV. The J1-*Bam*HI 4.0 kb fragment has been shown to contain the exons II (Figure 5.5 B). The exon I is not included in the isolated genomic clone, as already suggested by Southern blotting. Partial nucleotide sequences of the introns revealed no deviations from the consensus sequences for splice site in the 5' -donor and 3' -acceptor splice site and all intron-exon splice junctions followed the GT/AG rule.

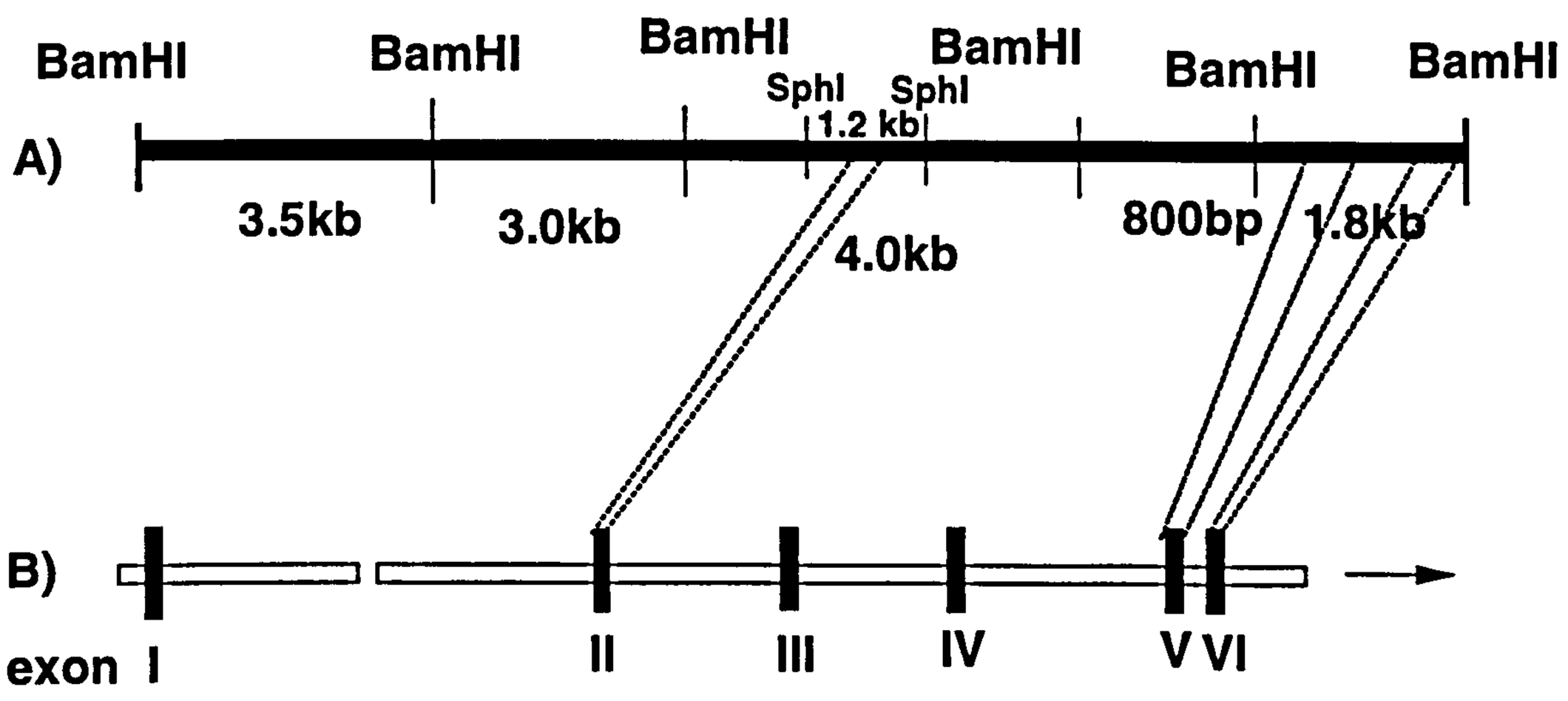


Figure 5.5. The analysis of the genomic organisation of the CRH-R1

Schematic representation of a *Bam*H I / *Sph* I restriction map of the isolated genomic clone (A), the deliniated exon-intron structure of the CRH-R1 gene. The partial nucleotide sequence confirmed that the clone contained exons II, V and VI. The black boxes represent the 3 exons identified. Introns are represented by double lines, the arrow represents genomic sequences that have not been analysed.

The exon I of CRH-R1gene contains a very GC-rich region (Figure 5.6). In order to confirm the clones contain the exon I of CRH-R1α gene, the J1-clone DNA was digested with restriction endonucleases, fractionated on agarose gel, and transferred to Hybond N filter membrane for Southern blot analysis. The filter was probed with [γ-^{32}P]dATP end-labelled oligonucleotide. The labelled oligonucleotide used a CRH-R1 specific primer (PE3 primer) 5'-CTTGACGAGACGGAGCTGCGG-3' that correspond to exon I. No positive hybridisation was found.

Further analysis was carried out using PCR 25ng of DNA was used in the PCR reaction, which was carried out for 20 cycles. The gene specific primers PE5 and PE3 correspond to -224 to -203 nt and +13 to +33 nt in exon I. The reaction was then analysed by electrophoresis on a 1.0% agarose gel. Sequence analysis of the PCR products did not correspond to any known DNA sequence. CRH-R1α (exon I) was not detectable in this clone.

Based on the sequence of the CRH-R1 exon I region shown in Figure 5.6, the *Sam* I enzyme can be used to cut this region at two sites and generate one product with 167 bp between the two cutting points. The cloned genomic DNA was digested with *Sam* I, fractionated on agarose gel, and visualised under UV light by ethidium bromide staining. The fragment obtained had a different size from that expected at the beginning of the experiment. Therefore, the result shows that the clone did not contain the exon I of CRH-R1α gene.

In order to obtain the transcriptional start site for CRH-R1 mRNA, we used the 5'Rapid amplification of cDNA ends (5'-RACE system Version 2.0 Gibco-BRL) technique using total RNA extracted from human myometrium. Briefly, cDNA was synthesised with CRF8 5'-GTCCACGGATGCGTTGCACTGC-3' primer and Superscript II Reverse Transcriptase. After RNase H treatment, the cDNA was purified with the Glass Max DNA isolation spin cartridge (Gibco-BRL). The purified cDNA was 5'tailed with polyC and then further amplified with an anchor primer provided by the kit and a gene-specific primer PE1. After primer extension and the

first round of PCR, the cDNAs obtained were amplified by the second round of PCR using a nested primer (PE3) and the kit's PCR anchor primer. The sequence of PE3 was 5'-CTTGACGAGACGGAGCTGCGG-3' which correspond to a region of exon I. The PCR products were cloned in the pGEM-T vector. Using this technique we were unable to find any cDNA extending upstream of exon I.

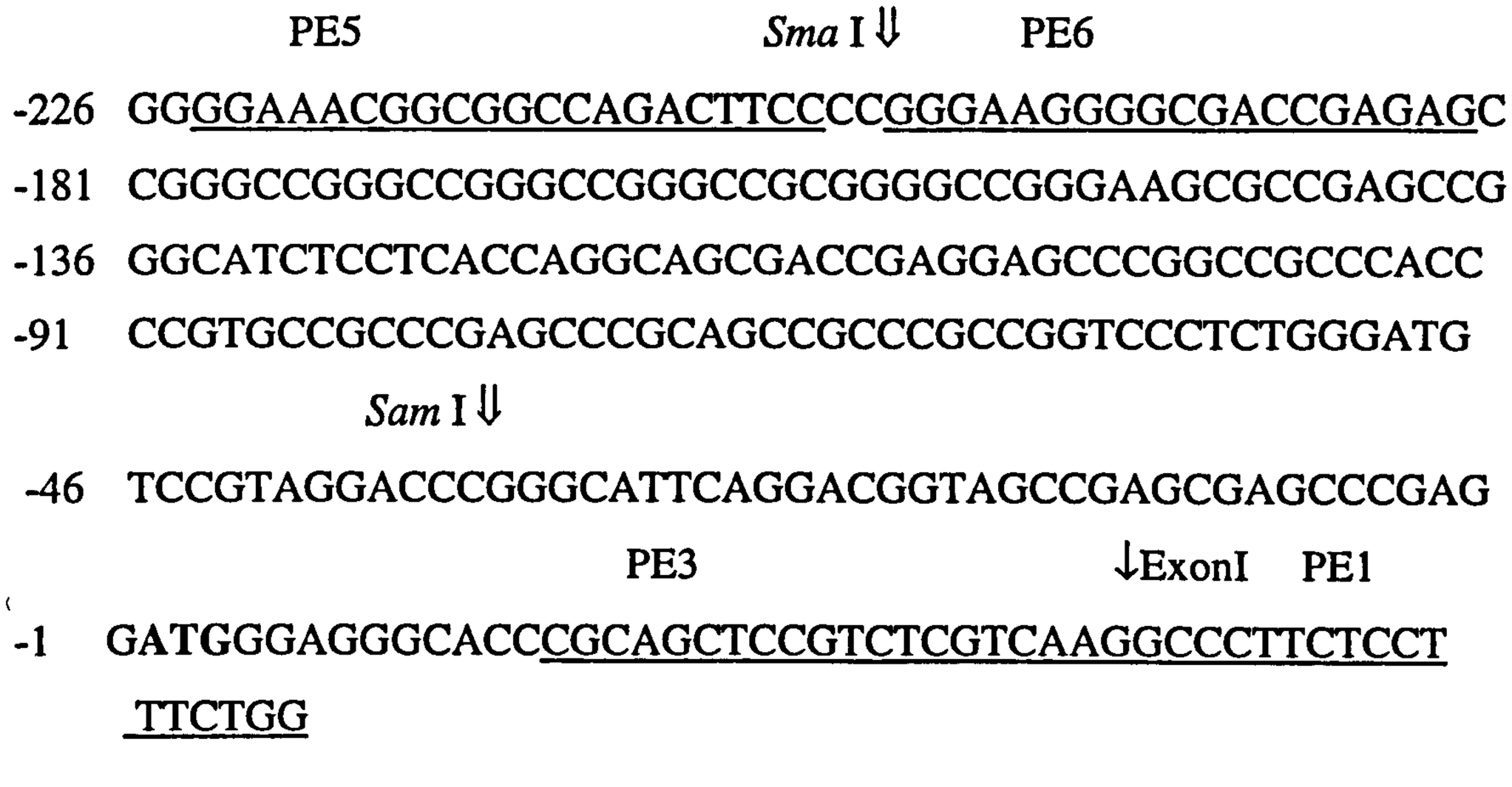


Figure 5.6. DNA sequence of the exon I region of the human CRH-R1 gene (↓) indicates intron position; (⇓) indicates *Sam* I restriction digest site. The underlined regions indicate PCR primers (PE5, PE6, PE3, and PE1). ATG translation initiation codon in exon I, is denoted as nt +1. Human CRH-R1 (GenBank/EMBL accession number x72304).

We also tried to detect the cDNA corresponding to the exon I, using either the primer PE5: 5'-GGAAACGGCGGCCAGACTTCC-3' or the primer PE3: 5'-CTTGACGAGACGGAGCTGCGG-3' (Figure 5.3), and total RNA extracted from human myometrium. We were unable to detect the cDNA corresponding to exon I, this may be due to the particular GC rich composition of this RNA segment.

Since the 5'-flanking region of the rat CRH-R1 gene was isolated and characterised (Tsai-Morris *et al.*, 1996), the present study attempted to use this information to identify the human homologue of CRH-R1 promoter region. We used the PCR technique to generate probe fragment for hybridisation to human library screening. PCR amplification of the 5'-flanking region of the rat CRH-R1 gene was performed using genomic DNA from rat brain tissue. The primer pair was based on the 5'-flanking region of the rat CRH-R1 sequence (GenBank/EMBL accession number: U53486). The PCR reaction was carried out for 35 cycles with an anneal temperature of 58°C. Sequence analysis of the PCR product did not correspond to a known DNA sequence of the rat CRH-R1 gene.

5.3 Discussion

At the time when this study started, no sequence of the promoter region of the human CRH-R1 had been reported. The initial attempt to screen the human placental genomic cosmid DNA library with a cDNA probe from exons I-V resulted in the isolation of only one clone. The clone was named J1, and was 30kb in length. Mapping by Southern analysis and sequencing confirmed that the clone contains exons II-VII of the CRH-R1. More specifically, the J1 *Bam*H I 1.8 fragment was subcloned and sequenced with various primers from both exon V and other up- and down-stream regions. Sequencing demonstrated that this fragment contained only exons V-VII. The J1-*Sph* I 1.2 kb was also subcloned, sequenced. The results showed that it contained exon II. The most upstream region identified in the library screen was the J1-*Bam*H I 4.0 fragment that was subcloned and sequenced with oligonucleotides to exon I (PE3 primer +13/33 bp and PE6 primer -183/202 bp). However, sequencing information obtained using either primer did not suggest that exon I resided more upstream of this fragment. Thus intron I of the human CRH-R1α is a large intron, and may span over 10 kb.

In order to isolate the 5'-upstream region of the CRH-R1 gene, we attempted to make a new probe with a 5'-end cDNA of the CRH-R1. The first approach was to look for the 5'-end cDNA of the CRH-R1 gene, using the 5'Rapid amplification of cDNA

ends (5'-RACE system Version 2.0 Gibco-BRL) technique and total RNA extracted from human myometrium. We failed to find any cDNA extending upstream of exon I. The results suggested that there was either a blocking of the primer extension by the GC rich region located at the 5'end of exon I, or the GC rich sequences that possess strong secondary structure that resists denaturation and prevents primer annealing.

We also tried to detect the cDNA corresponding to the exon I of the human CRH-R1. RT-PCR was amplified using a pair of gene specific primers for the exon I, but we were unable to detect the cDNA corresponding to the exon I, which may be because of the particularly GC rich composition of this RNA segment. This could also be due to the GC rich region, located at the exon 1, which could have blocked the primer extension step.

It has been reported that the 5'-flanking region of the rat CRH-R1 gene was isolated and characterised (Tsai-Morris *et al.*, 1996). This gave me some knowledge to do with in designing the PCR primer to generate a probe. PCR was performed using rat brain DNA as template and amplified to the promoter region of the rat CRH-R1 gene. But the promoter was not found in which may be due to the high GC content of this region. The PCR products were still failed to yield any 5'-flanking region of the rat CRH-R1 DNA.

Although the promoter of the human CRH-R1 was not found from the above work, some findings may be interesting. The intron-exon junctions conform entirely to the consensus spice site sequence of 5' donor (GT) and 3' acceptor (AG) (Mount, 1982). These are in consistence with the view that the sequence region (exons II-VII of CRH-R1) is the obligatorily conserved nucleotides, which include GT in the 5'-donor and AG in the 3'-acceptor region. Another finding is that the exons II-VII of CRH-R1 sequences matched the previously published human CRH-R1 cDNA sequences (GenBank accession no. AF039510 to AF039516).

Chapter 6 Screening of a human genomic cosmid library (CRH-R2)

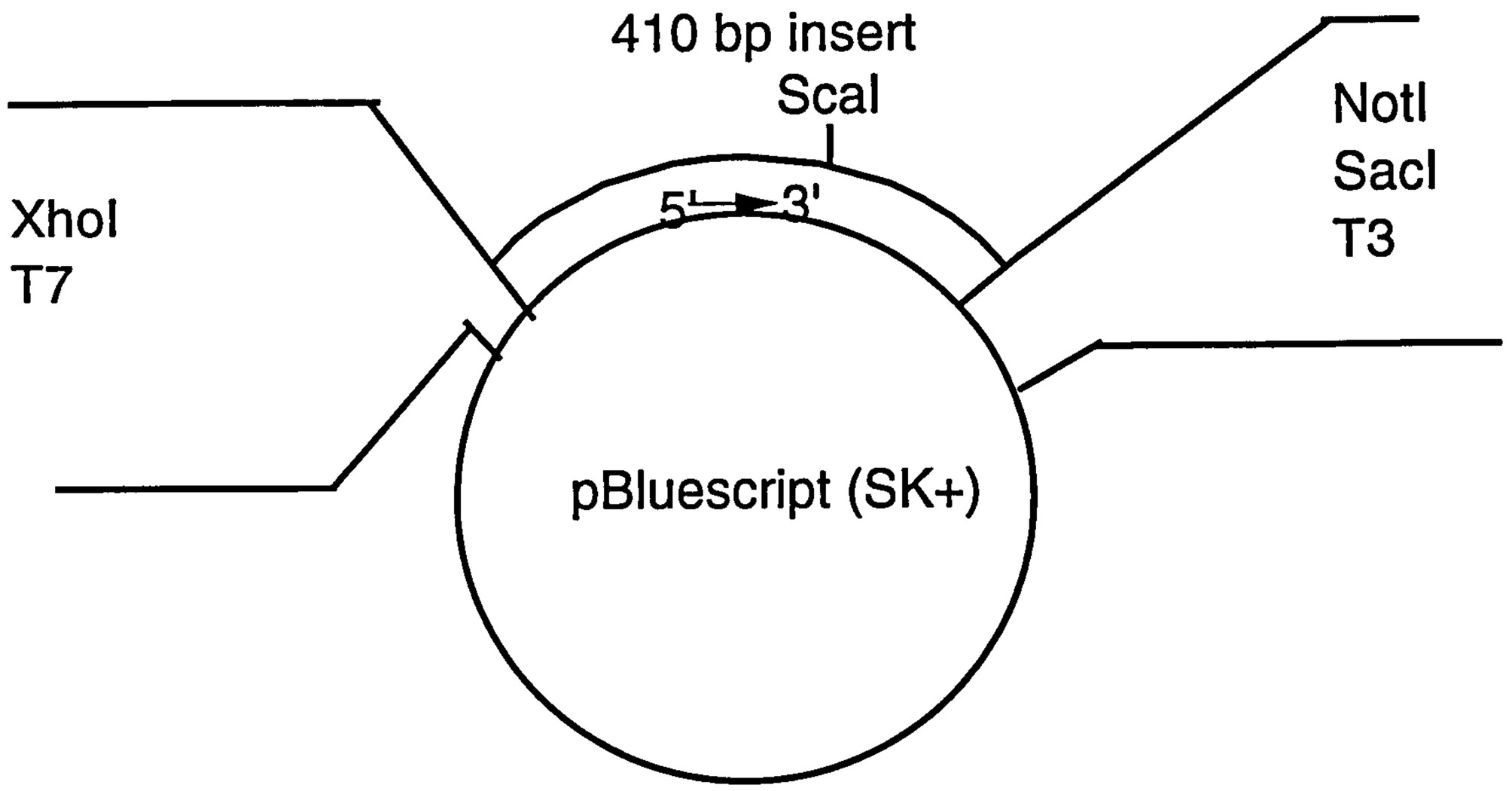
6.1 Introduction

The gene coding for human CRH-R2α consists of at least 12 exons and spans approximately 30 kilobases, with the protein-coding sequence interrupted by 11 introns (Liaw *et al.*, 1996). The gene has been mapped to the chromosome 7q14-15 (Andrews *et al.*, 1998). The genomic structure of a significant portion of the human CRH-R2 gene has been described, but this did not extend the 5'of the CRH-R2α alternative exon. Neither the 5'-flanking region nor the transcriptional initiation site has been defined. Without upstream sequence data, the precise mechanism of the receptor's regulation remains unknown. To understand the molecular mechanisms that direct the expression of the gene encoding the CRH-R2, a genomic clone of the CRH-R2 was isolated and analysed.

6.2 Screening and its results analysis

6.2.1 Isolation of genomic DNA clone containing 5'-flanking region CRH-R2α gene

Screening of ≈ 500,000 clones of a human genomic cosmid library with a ^{32}P-labeled DNA containing nucleotides 1-197 bp of CRH-R2α cDNA probe (Figure 6.1) resulted in the isolation of one positive clone. Positive clones from the second or third round screens were picked, and minipreped using QIAprep spin plasmid kit (Qiagen). The clone was named Jp 1.



Figur 6.1 Human CRH-R2 probe inserted into the pBluescript (SK+) cloning vector. The plasmids was digested with *Xho* I and *Sca* I, and purified from a 1.0% agarose gel using "QIAquick Gel Extraction kit" (Qiagen).

Phage DNA from this positive clone was prepared and digested with *Not* I restriction enzyme to release the insert. Restriction digestion of the genomic clone with *Sst* I released seven fragments of approximately 5kb, 4kb, 3kb, 1.8kb, 1.2kb, 1.0kb and 650bp. The DNA from this positive clone was digested with *Sph* I which released nine fragments of approximately 5kb, 3.5kb, 3.2kb, 2.5kb, 1.8kb, 1.2kb, 700bp, 600bp, and 450bp.

The genomic clone DNA was digested, fractionated on agarose gel and transferred to Hybond-N filter membrane for Southern blot analysis. The filter was probed with ^{32}P-labeled cDNA corresponding to the region of human CRH-R2α cDNA as indicated in exon I- II. This was hybridised to two restriction fragments, a JP1-*Sst* I 4.0 kb

fragment and a JP1-*Sph* I 3.5 kb fragment (Figure 6.2 A). In addition, the JP1-*Sst* I 4.0 kb fragment was digested again with *Nco* I and *Kpn* I and probed with the radiolabelled cDNA. This identified 2 positive fragments a JP1-*Kpn* I 1.7 kb and a JP1-*Nco* I 2.5 kb. (Figure 6.2 B) These fragments were subcloned and sequenced. The nucleic acid sequence of the fragments revealed that, both Jp1-*Kpn* I 1.7 kb and *Nco* I 2.5 kb, contained exon I (103 bp), intron I (137 bp) and exon II (126 bp). Comparing the genomic DNA with the cDNA sequence, the location of intron I was identified.

6.2.2 Isolation of genomic DNA clone containing 5'-flanking region CRH-R2γ gene

The CRH-R2 has three spliced variants, CRH-R2α, CRH-R2β, CRH-R2γ. Each of the products has different 5'-exon splicing and distinct tissue distributions. The three splice variants appear to diverge at the corresponding position of human CRH-R2 where the first intron occurs. The three first exons are alternatively spliced on to the first common exon (CRH-R2α exon II), followed by a second common exon. Based on the results of the genomic screen described above, the CRH-R2α exon lies downstream of the CRH-R2γ and CRH-R2β exons. In order to confirm this the JP1-clone DNA was digested with restriction endonucleases, fractionated on agarose gel, and transferred to Hybond N filter membrane for Southern blot analysis. The filter was probed with [γ-^{32}P]dATP end-labelled oligonucleotide designed from the CRH-R2γ exon I (5'-CTGTGCTCAAGCAATCTGCCTAC-3'). This probe hybridised to a unique JP-1 *Sph* I 3.2 kb fragment (Fig 6.3) which was subcloned and sequenced.

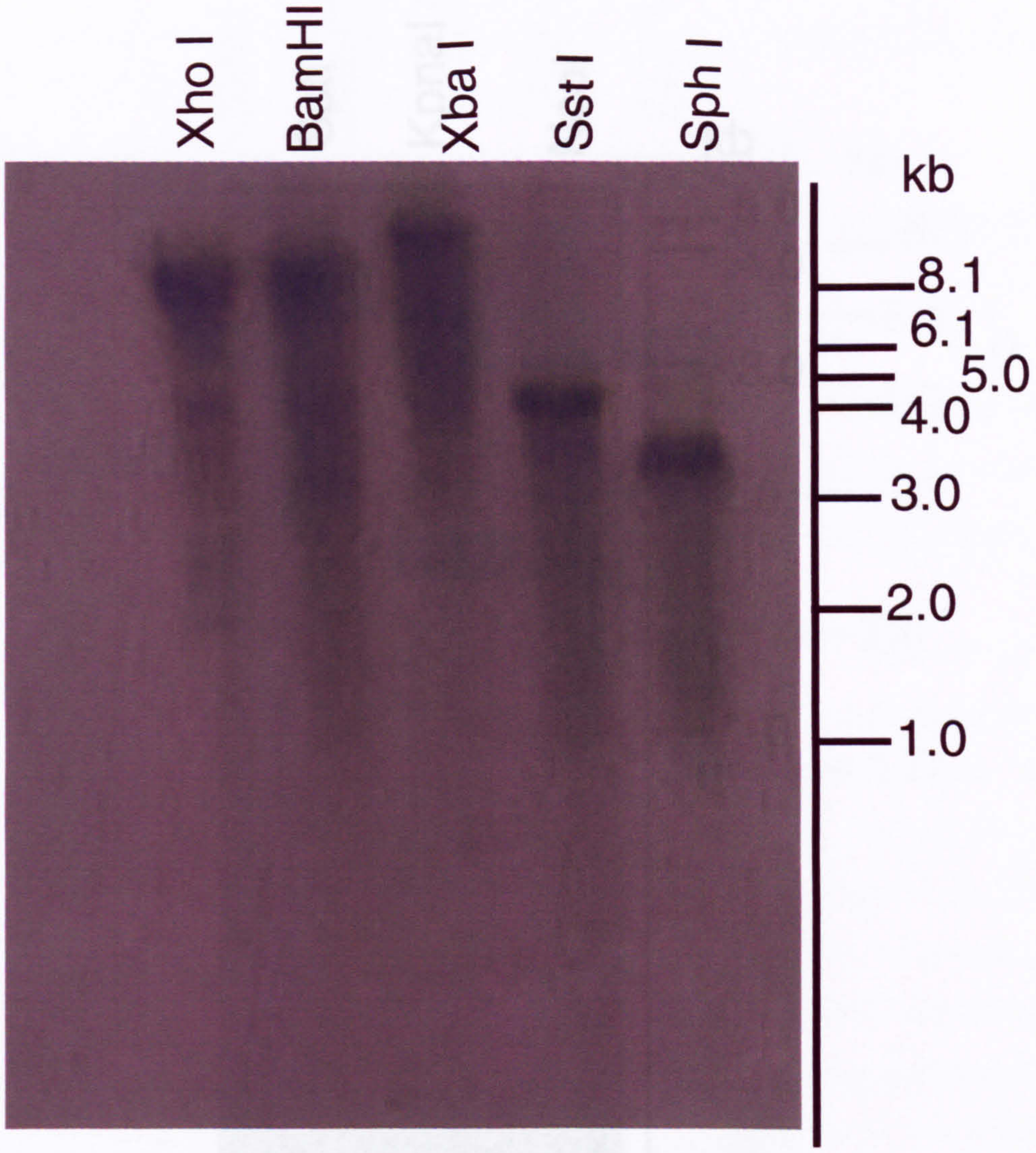


Figure 6.2 (A) Southern blot analysis of clone containing 5'-Flanking region CRH-R2α

Each lane contains the clone DNA which was digested each with *Sph* I, *Sst* I, *Xho* I, *Bam*H I *and Xba* I, respectively. After electrophoresis and blotting onto Hybond N, different strips of the same run were hybridised with an exons I-II probe of the CRH-R2α gene. This hybridised to two restrictions fragments a *Sph* I 3.5 kb fragment and a *Sst* I 4.0 kb fragment.

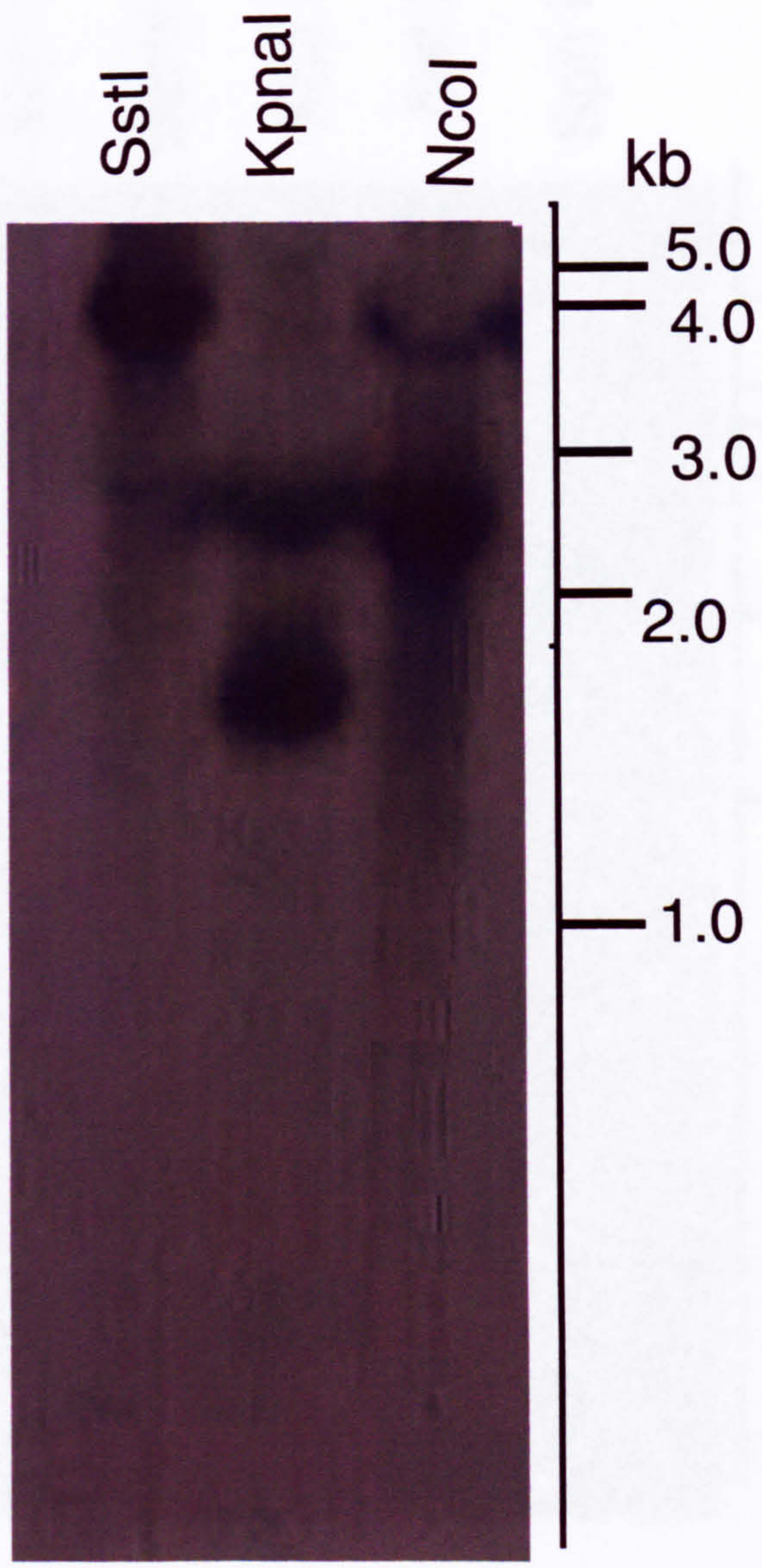


Figure 6.2 (B) Southern blot analysis of the clone containing 5'-Flanking region CRH-R2α

The *Sst* I 4.0 kb fragment was digested again with *Nco* I and *Kpn* I., fractionated on agarose gel and transferred to Hybond-N. Filter was probed with the radiolabeled cDNA from corresponding to exon I- II of the human CRH-R2α. This identified 2 positive fragments a JP1-*Kpn* I 1.7 kb and a JP1-*Nco* I 2.5 kb.

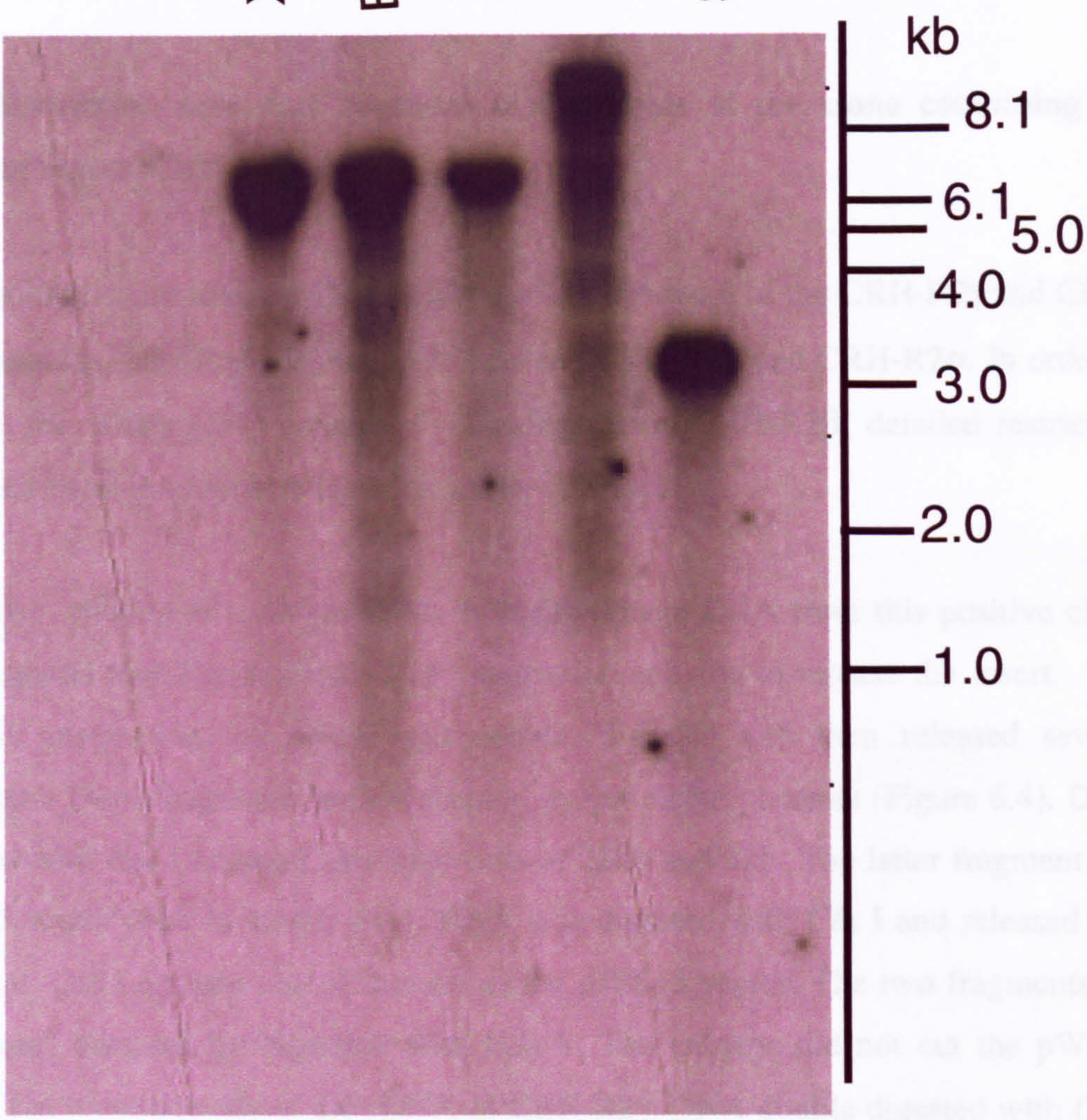


Figure 6.3 Southern blot analysis of the clone containing 5'-Flanking region of CRH-R2γ

The isolated clone DNA was digested with *Sph* I, *Sst* I, *Xho* I, *Bam*H I *and Xba* I, respectively. After electrophoresis and blotting onto a nylon membrane (Hybond N), different strips of the same run were hybridised with an exon-I probe of the CRH-R2γ gene. A unique *Sph* I 3.2 kb fragment signal was found.

The nucleotide sequence of the genomic DNA fragment showed that it contained the CRH-R2γ exon I (227 bp) and the CRH-R2β intron Ib (205 bp). These results indicate that the CRH-R2β exon lies upstream of the CRH-R2γ and CRH-R2α. The genomic order of the alternative exons is CRH-R 2β-2γ-2α.

6.2.3 Restriction map and Southern blot analysis of the clone containing 5'-flanking region CRH-R2β

Our results indicate that the CRH-R2β exon lies upstream of the CRH-R2γ and CRH-R2α, and also, the CRH-R2γ exon lies between CRH-R2β and CRH-R2α. In order to confirm that clone (JP1) contains 5'-flanking region CRH-R2β, detailed restriction maps and Southern blot hybridisation were performed.

The clone, entitled JP1, was ≈ 40 kb in length. Phage DNA from this positive clone was prepared and digested with *Not* I restriction enzyme to release the insert. The genomic clone was, by single and double, digested and then released several fragments. These fragments were separated by gel electrophoresis (Figure 6.4). DNA digested with *Not* I released two fragments of 23kb and 8kb. The latter fragment is a pWE15 vector (8kb in length size). DNA was digested with *Cla* I and released one fragment. *Cla* I cut one site of the end of the pWE15 vector. The two fragments (as same size) obtained by digestion with *Nde* I.. The enzyme did not cut the pWE15 vector, but it cut one site of the cDNA of CRH-R2γ. DNA double digested with *Cla* I and *Nde* I released three fragments of approximate 8kb, 19kb, and 21kb. The double digestion with *Not* I and *Nde* I obtained the three fragments. The double digested with *Cla* I, *Nde* I or *Not* I, *Nde* I had it of size fragments. When *Cla* I cut one end of the pWE15 vector, *Nde* I cut another end of the vector, it released one fragment of approximate 8-8.5kb (possible only vector). If this is the case, the position of the CRH-R2γ is likely at one end of the insert. This clone did not include the exon I of the CRH-R2β. The nucleotide sequence of the genomic DNA fragment revealed that the clone did not include CRH-R2β exon I.

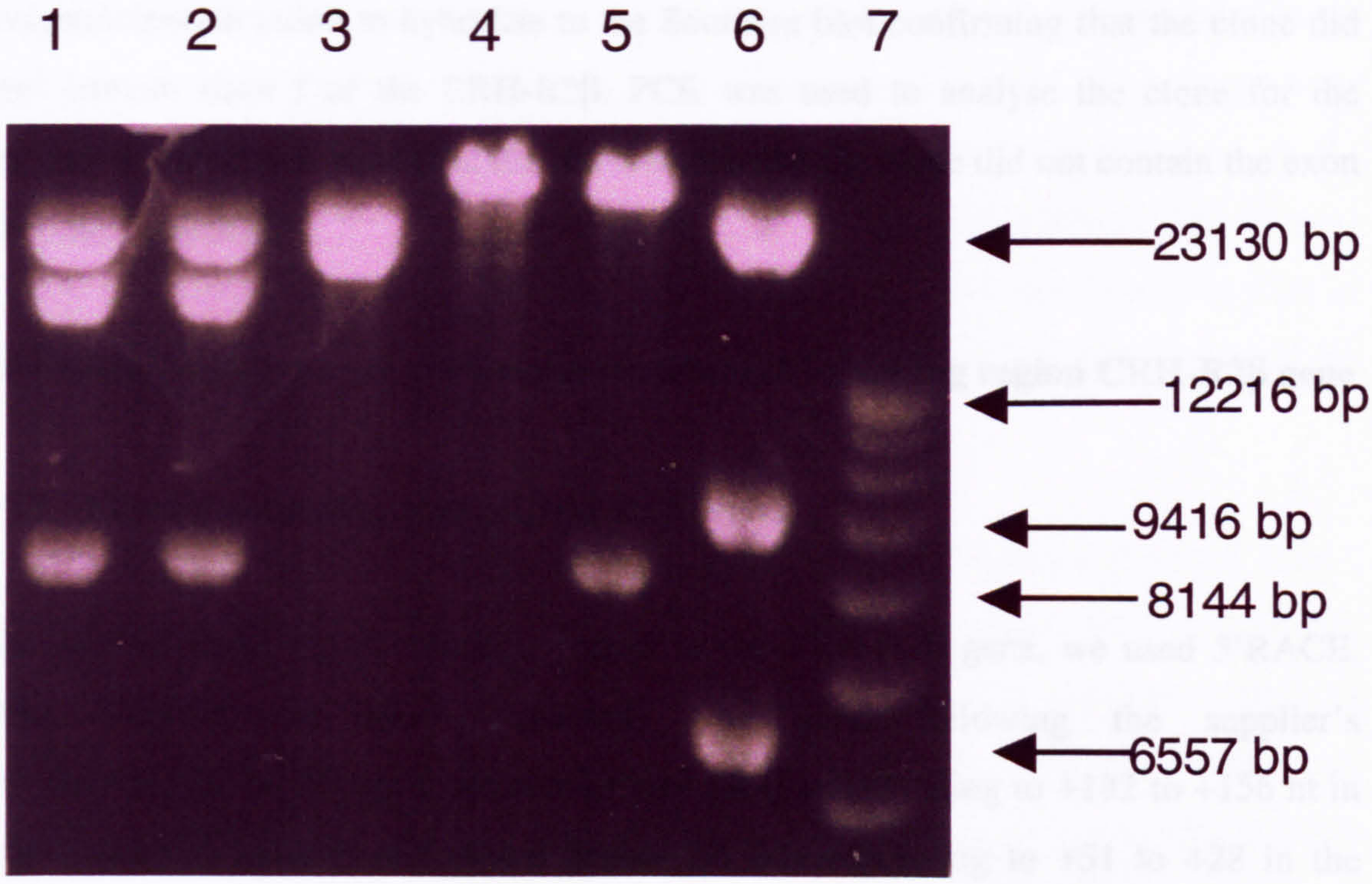


Figure 6.4 Restriction map analysis of the clone containing 5'-flanking region CRH-R2β

Lane 7 is 1Kb DNA ladders. Lane 6 *Hind* III fragments of bacteriophage λ were included as size markers. The clone was digested with single and double restriction enzymes give distributions of digestion products. Lane 5, the comid clone digested with *Not* I released two fragments of 23 kb (over) and 8 kb, respectively in sizes. Lane 4 the cosmid clone was digested with *Cla* I released one fragment. Lane 3 the two fragments (with same size) were obtained by digestion with *Nde* I. Lane 2 the cosmid clone double digested with *Cla* I and *Nde* I released three fragments of 8kb, 19kb, and 21 kb, respectively. Lane 1 the cosmid clone double digested with *Not* I and *Nde* I released three fragments of 8kb, 19kb, and 21 kb, respectively.

We also examined whether the clone contained the exon I of the CRH-R2β gene, using the synthetic oligonucleotide S1 (5'-TCCAGTCCCTAACCCCAGCC-3'), corresponding to a sequence located in exon I. The [γ-^{32}P]dATP end-labelled oligonucleotide failed to hybridise to the Southern blot confirming that the clone did not contain exon I of the CRH-R2β. PCR was used to analyse the clone for the presence and size of insert, but the result shows that the clone did not contain the exon I of the CRH-R2β.

6.3 Isolation of genomic DNA clone containing 5'-flanking region CRH-R2β gene

6.3.1 Library screening with a CRH-R2β probe

In order to isolate the 5'-flanking region of the CRH-R2β gene, we used 5'RACE. The 5'RACE kit from (Clontech) was used, following the supplier's recommendations. The gene specific primer 1A (corresponding to +182 to +156 nt in the CRH-R2β exon-I) and nested primer 2A (corresponding to +51 to +28 in the CRH-R2β exon-I) were used for the first and second amplifications, respectively, with plus anchor primer AP1 supplied by Clontech. The resultant products were cloned into the pGEM-T cloning vector and the clones were sequenced. The sequence resulted in one DNA fragment of apparently 255 bp in the human hippocampus sample. The 5'RACE derived from cDNA clones included the previously reported 30 nt 5'-UTR sequence upstream of the CRH-R2β open reading frame as well as an additional 171 nt of 5'-UTR sequence.

This clone was selected for use as a probe in the screening of a genomic cosmid DNA library. An amplified human placental genomic cosmid DNA library (Clontch) was employed to isolate clones containing the 5'-flanking of the CRH-R2β gene. Plaque lift of 1X10^6 recombinants were hybridised with a 255-base pair ^{32}P-labelled (Oligolabelling kit, Pharmacia Biotech) fragment derived from the 5'-end (+54 to -201 bp, ATG translation initiation codon in exon I, which is denoted as nt +1) of a

human CRH-R2β clone. 1,000,000-phage clones were screened but no positive clones were isolated.

6.3.2 Genomic walking

Since the introduction of the polymerase chain reaction, several PCR-based cloning methods to isolate genomic DNA have been devised. Several PCR-based methods have been used for genomic walking. These include inverse PCR (Ochman *et al.*, 1988), ligation-mediated PCR (Mueller and Wold, 1991), Biotin-RAGE (Bloomquist *et al.*, 1992), one-armed PCR (Macrae and Brenner, 1994) and rapid amplification of genomic ends (RAGE) PCR (Buckler *et al.*, 1991). The last method has been used for the genomic walking in the project.

A CRH-R2β specific primer was designed, based on the published CRH-R2β cDNA (Valdenaire *et al.*, 1997) and 5'RACE-derived cDNA sequence. Human genomic DNA (from peripheral blood leukocytes) was first incubated with a restriction enzyme (*Nco* I) that recognises a site within the 5' end of a gene, followed by denaturation and polyadenylation of its free 3'ends with terminal transferase. The modified DNA was then used as a template for PCR using a gene-specific primer complementary to a sequence in the 3' end of its cDNA (located 51-28 bp downstream of the ATG translation initiation site) and anchored deoxyoligothymidine primer. A second round of PCR was then performed with a second, nested gene-specific primer (located 121-146 bp upstream of the ATG translation initiation site) and the anchor sequence primer (see Chapter 2.2 18.2).

The final PCR products were size-selected by gel electrophoresis, (Figure 6.5) and subcloned into the pGEM-T vector (Promega). Several clones were selected and sequenced in both directions. The insert DNA was cycle-sequenced using an ABI 373A (Applied Biosystems) automated fluorescent sequencer.

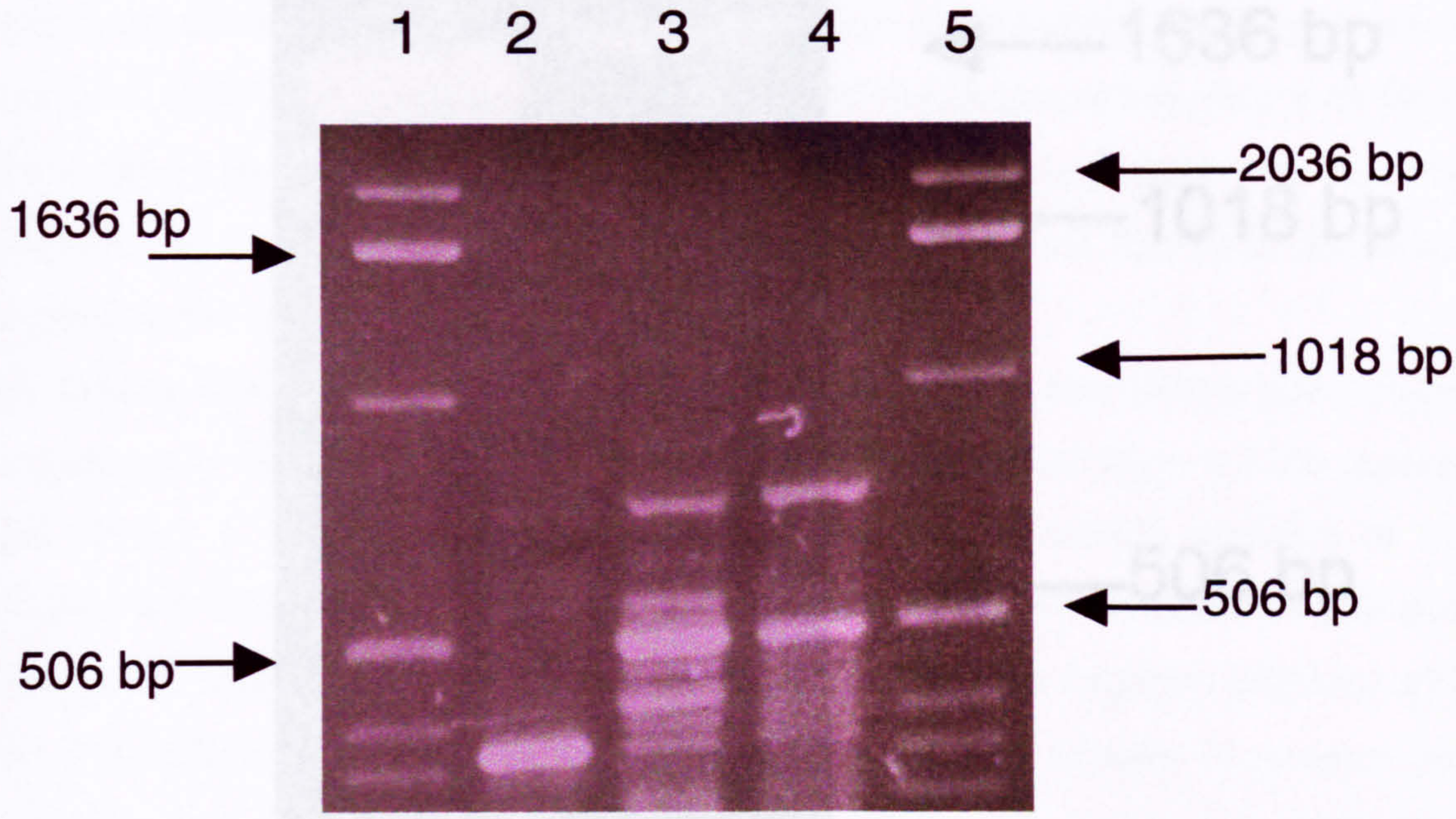


Figure 6.5 The 5'-flanking region of the CRH-R2 gene by the PCR based genomic walking method.

A nested PCR was used to amplify the 5'-flanking region of the CRH-R2 gene. The DNA was digested with *Bam*H I (Lane 2) or *Nco* I (Lane 3 and 4 are samples with different concentrations). Lane 4 shows the PCR amplification of a 734 bp fragment, containing the 5'-flanking region of the CRH-R2 gene. Lane 1 and 5 are DNA size markers.

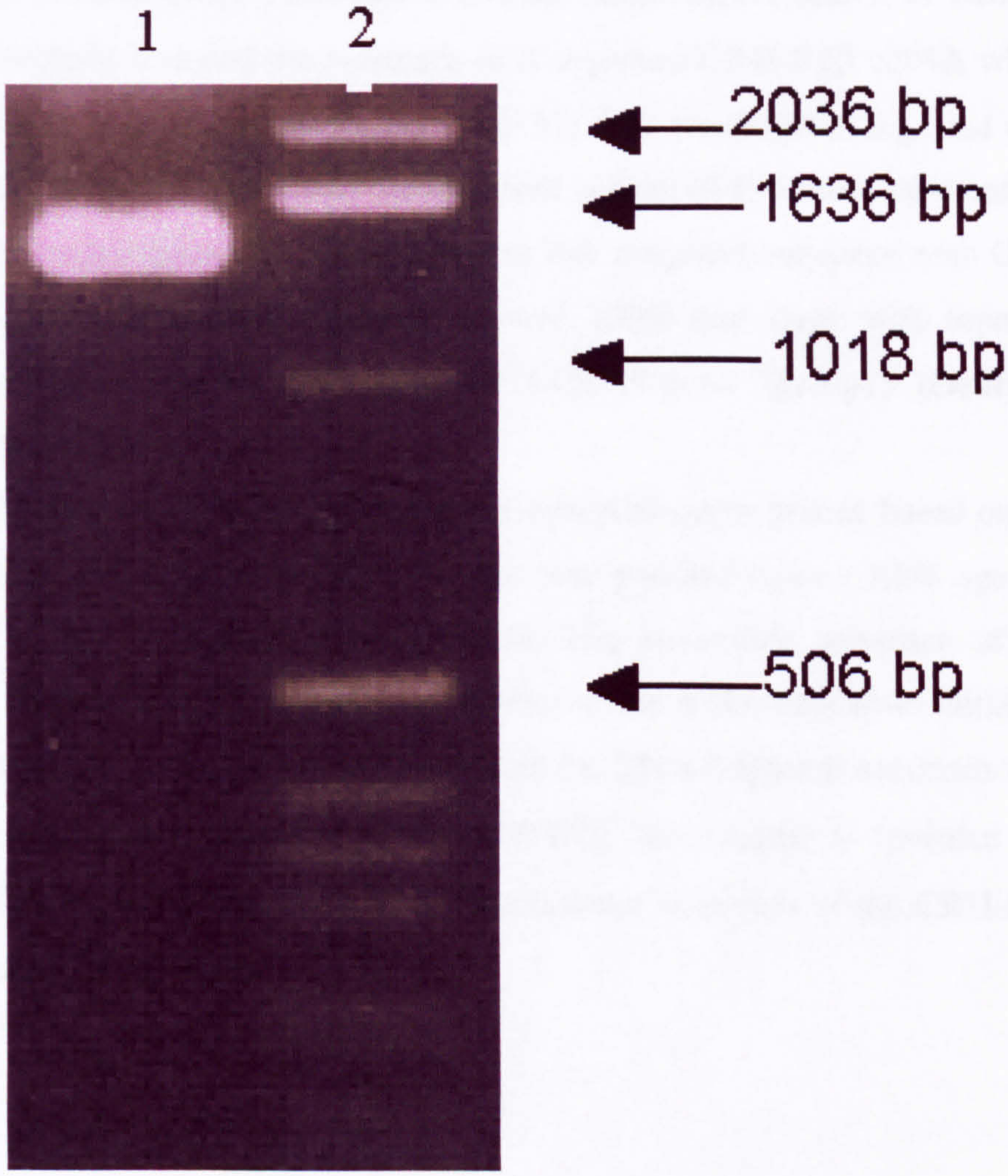


Figure 6.6 PCR amplification further upstream of the 5'-flanking region of the CRH-R2 gene.

Lane 1 shows the PCR amplification of a 1440 bp fragment, which contained 1393 bp upstream of the ATG translation initiation site and an 51 bp downstream sequence. Lane 2 are DNA size markers.

6.3.3 Database sequence analysis

The sequence data was analysed using Blast Nucleic Acid Database Searches from the National Centre Biotechnology Information (NCBI). Blast server search of existing sequences in GenBank revealed the existence of a sequence CRH-R2β cDNA which has been published (GenBank accession No. Y10151). The result (DNA digested with *Nco* I restriction enzyme) revealed that the fragment contained 835 bases upstream of the ATG translation initiation site. Comparison of this fragment sequence with Gene Bank data base revealed that the sequence showed 100% homology with reported sequence of the Homo Sapiens PAC clone DJ1143H19 from 7p14-p15 (GenBank accession No. AC004976).

In order to obtain the region further upstream, we designed a new primer based on the sequences of the above results. The PCR product was purified from a 1.0% agarose gel (Figure 6.7), then subcloned and sequenced. The nucleotide sequence of the fragment revealed that a 1393 bp upstream sequence of the ATG translation initiation site and 51 bp downstream sequence. Alignment of the DNA fragment sequence with the 5'RACE derived cDNA sequence of the CRH-R2β (see Chapter 8) revealed that the DNA fragment included the 226 nt 5'-UTR sequence upstream of the CRH-R2β open reading frame (Figure 6.8).

-1393 AGACCTGTGAGATTTCACAGGAGAGGGAGGTGGGACCTGGAAC

Bgl I

-1350AGCCAGGAAGGCTGCCTGGAGGAAGGGGCTGAGGGTGGGGAGGAAGCACT

-1300CAAGGACAGAAAAGTGTCCCATGTCTTAGATAGGGAGACCGGAAGGAGCA

-1250 CAGAGTGTCTGAGTTACTTAAATCGGGACTGAGTTTTCCAAGGTCTCTCT

-1200 CCACCATGCCTGCATCTCCAGGAGAAAGTTTGGTGCCCCACTAATCACAT

-1150 TCCTCACATGCTTTCCCCTCAGATCCTAACAACAAGCAAGGTGCAACTGT

Hind III *Dra* I

-1100 GGCTTAAGCTTTTAAACCCCCTGCACCAGGCCCACCTGTCAGGAACAGCT

-1050CCTAAGGAGGCTTCTGAGGCTCTGGAAGAACAGACCTCACCTCATGAGCC

-1000CTGACCCAGCCCCAGGTTGAACCTGTCCCCGCCAAGAACAGCACTGCTCA

EcoR V *Bam*H I

-950 GCCACCTGATATCAGGATCCCTAACCATTTCCTGAAATGGAACCTGGGCC

-900 CCATACAGACAGACACCTGGGCCCCATCCCAAGCTACTTCATTGGAATCT

Nco I

-850 CTGGGTGAGCCCTGCCATGGCTGTATTTGCTGTCGGTTGGGTTTTTTTTT

-800 TTTTTTTTTGAAAGCATTTTATTTTGATAGAATTTCAAACGTACAGAAAA

-750 GTTGGAAGAATGATCCAAAGCATATCCATATACCCCTTTTCCAGATTTAC

-700 CAACAATTTACATTTTGCCCCATCTGCTCTCTTGTTCTATCTATGTATCT

-650 ATCTGAGTTTATTTCTGAACCATTTGAAAGTATGTTGGAAACATGGTGCC

-600 CCTTTAATCCTAAATGTGTATTTTCCAAGAATGAGAACATTCTCTTACCT

Sca I

-550 ACCCACAGTACAGTATCAGAATCAGCAGATGTAGCATTGATACAGTACTA

-500 TCTAACCCAGCATCCACACTCCCATCTCAGTCATCCCAGTAGGCATTAGC

-450ACCTGTTAAAAGCTCCTCAGATAGGTGTTAGGCAGTGCTGGTTCAGCTGA

-400GACCCAGGGCAGCCCAACCCTGTGCTGAAGCAGGGACCACAGCCTAGGCC

Xma I

-350 CCTCCTGCCTGTCTCAATTCCTCTCTCCCACCTGGGGAGCTACAGCCCCG

Pst I

-300 GGCAATGCTGGCAGCCGATAGAAGGCATTAACCCCTTCTCTGCAGGCTGG

-250 ATGGAGAGCATAGCCCCTTGGCTGGCTGGGTAGCAGCAGCTGTCTGGGAC

⇑ 2β

-200 CAGCTATAAATAACCTGGAGGCTGTGGGCTGTGGGGCGGGGTCACCACTG

-150 CCCTGTGGAAGAAGTGGGTGGAGAATACCTCCTCATGCCCAAGGCCTTAC

2nd : 2A

Bal I

-100 CTTGGCCTCCCCTCCAGGGCCGGGTGGGGTGGGGCTGGCCAGGGTGTGAC

-50 CACCGTGCTGGGCAGCAGGCTCCAGTCCCTAACCCCCAGCCACTACTGGC

V ↑

+ 1 ATGAGGGGTCCCTCAGGGCCCCCAGGCCTCCTCTACGTCCCACACCTCCTC

1st: 1A

Figure 6.7 Nucleotide sequence of the 1.3 kb 5'-flanking region of the CRH-R2 gene.

The translation start site is indicated by the mark defined as + 1. Nucleotide numbers are marked on the left side. The primers used for PCR (1A and 2A) are underlined. (↑) marks the location of the known cDNA previously reported by

Valdenaire *et al.* (⇑) represents the transcription start site identified by 5'RACE with CRH-R2β primers.

6.4 Discussion

From the description of above sections, we can see that screening of a human genomic cosmid library has been carried out and resulted in the isolation of one positive clone Jp 1. Mapping by Southern hybridisation and sequencing analysis confirmed that the clone contains 5'-flanking CRH-R2γ and CRH-2α. More specifically, the Jp1-*Sst* I 4.0 kb fragment was subclond and sequenced with various primers from both exon V and other up- and downstream regions. Sequencing analysis demonstrated that this fragment contained CRH-R2α exon I (103 bp), CRH-R2α intron I (137 bp) and exon II (126 bp). Comparing the genomic DNA with cDNA sequence of the CRH-R2, the location of intron Iα was identified. The JP-1 *Sph* I 3.2 kb fragment was subcloned and sequenced. The nucleotide sequence of the genomic DNA fragment revealed that the clone included CRH-R2γ exon I (227 bp), CRH-R2γ intron I (2831 bp) and CRH-R2β intron Ib (205 bp). Unfortunately, screening of a human genomic library with a probe specific for the 5'-region of the CRH-R2β was unsuccessful.

We have cloned the 5'-flanking region of the CRH-R2β gene by the PCR based genomic walking method. A CRH-R2β specific primer was designed based on the published CRH-R2β cDNA (Valdenaire *et al.*, 1997) and the 5'RACE-derived cDNA sequence which were then amplified, cloned and sequenced. The nucleic acid sequence of the fragment revealed that 1393 nt sequence upstream of the CRH-R2β open reading frame and an 51 nt downstream sequence. The 1393 nt fragment included the previously reported 30 nt 5'-UTR sequence upstream of the CRH-R2β open reading frame. Alignment of the DNA fragment sequence with the 5'RACE derived cDNA sequence of the CRH-R2β (see Chapter 8) revealed that the DNA fragment included the 226 nt 5'-UTR sequence upstream of the CRH-R2β open reading frame. Comparison of this fragment sequence with GenBank data base

revealed that the fragment corresponded to the Homo Sapiens PAC clone DJ1143H19 from 7p14-p15 (GenBank No: AC004976). This PAC clone includes the 5' exon of the CRH-R2 type.

The 5'-untranslated region extends 226 bp (5'RACE) upstream of the initiator codon and is GC-rich (61%). Some members of the secretin and glucagon receptor family have multiple untranslated (non-coding region) exons in the 5'-untranslated region. The parathyroid hormone (PTH)/PTH-related peptide (PTHrP) receptor gene has three untranslated exons in the 5'-untranslated region (Manen *et al.*, 1998). The pituitary adenylate cyclase-activating polypeptide (PACAP) receptor gene has four untranslated exons in the 5'-untranslated region (Chatterjee *et al.*, 1997). Alternative splicing of 5'-untranslated regions of genes for other G protein receptors, human A1 adenosine, PTH/PTHrP, and PACAP receptors, seems to be related to the tissue-specific expression of their transcripts. However, this does not appear to be the case for the CRH-R2. 5'RACE (with CRH-R2 primers) analysis which identified four transcriptional start sites located at 226, 201, 97, and 12 bp upstream of the translation start site of the CRH-R2β. This is a hypothesis or prediction only and further experiments may be required to provide more solid evidence. On the other hand, it is probable that there is a single start site at 226 bp. It is also more likely that the 12, 97, and 201 bp are incomplete transcripts (see Charter 8). The results suggest that there may be a single promoter regulating the expression of all (CRH-R2α, 2β, and 2γ) subtypes. The promoter of CRH-R2 lies upstream of the 5'-end of the CRH-R2β.

Multiple promoter regions have been described for PTH/PTHrP and glucagon receptor genes. The PTH/PTHrP receptor gene has been shown to be controlled by two promoters, one of which is not GC-rich and is only active in kidney (McCuaig *et al.*, 1995). We cannot rule out the existence of multiple promoters in the CRH-R2 gene.

Chapter 7 Analysis of the human 5'end CRH-R2 genomic DNA structure

7.1 Introduction

The CRH-R2 gene has been mapped on to the chromosome 7q14-15 (Andrews *et al.*, 1998). The gene coding for CRH-R2α consists of at least 12 exons and spans approximately 30 kilobases, with the protein-coding sequence interrupted by 11 introns. The exons vary in length from 42-154 bp and are separated by introns that vary in size from 130 bp to 14.5 kb. The N-terminal extracellular domain of the receptor is encoded by the first three exons and a part of the fourth exon. Transmembrane domains I, II, III, and VI are entirely within exons 4, 5, 6, and 10, respectively, while transmembrane domains IV, V, and VII are each interrupted by a single intron. When appropriately spliced, it should yield a message with 1233 bp of open reading frame, with no difference from the available corresponding cDNA sequence. All exon/intron junctions are confirmed with the consensus sequences, with exons flanked by AG and GT at the 5' and 3'end, respectively (Liaw *et al.*, 1996).

Molecular cloning techniques have identified three alternatively spliced forms of the CRH-R2 differing in their N-terminal domains, identified as CRH-R2α, CRH-R2β and CRH-R2γ. The genomic structure of the human CRH-R2α gene was described, but this did not extend 5'end of the CRH-R2 alternative exon. Our results described in Chapter 8 suggest that the CRH-R2 genomic order of the alternative exons is 2β-2γ-2α. The mature receptor subtype mRNA appears to be generated via 5' exon differential splicing. The detailed analysis of the 5' end of the CRH-R2 genomic DNA structure is given in this Chapter.

7.2 Sequence analysis of the genomic region between the CRH-R2γ and CRH-R2α exons

Screening of a human genomic cosmid library with a ^{32}P-labeled cDNA containing nucleotides 1-197 of CRH-R2α cDNA probe resulted in the isolation of one positive

clone, Jp 1. Mapping by Southern analysis and sequencing confirmed that the clones contained 5'flanking CRH-R2γ and CRH-2α. (As discussed as Chapter 6).

In order to analyse the 5'end CRH-R2 of genomic region with sequences, two of the four following sequencing strategies were chosen:

(1) Primer walking.

If the segment of DNA to be sequenced is small, a primer that binds to a known sequence in the flanking DNA can be used to initiate a sequence run across the region of interest. A second sequencing run over the same region from the opposite direction is highly recommended, since factors that affect the sequence quality on one strand are often not present in the opposite strand. If the length of the DNA to be sequenced exceeds the length of a single sequencing run, sequence data from the far end of the first run can be used to prepare the primer of the far end of a second run into the region of interest. This process, called primer walking, can be repeated many times to sequence extensive tracts of DNA. The major advantages of primer walking are that no subcloning is required, and the location and direction of each sequencing run are known. In order to sequence the genomic region between the CRH-R2α and CRH-R2γ exons, this method was employed.

(2) Nested deletion.

Nested deletion strategies work best for DNA segments in the 5-10 kb range and require some prior knowledge of the restriction sites within the target DNA. Nested deletion with exonuclease III requires a unique blunt-end or 5'overhang restriction site adjacent to the region to be deleted, and a unique 3'overhang site distal to the deleted region to protect the vector DNA. When these sites are exposed after cleavage, exonuclease III will enter the DNA at the 5'overhang site and begin to digest or strand in the 3'→5' direction. The enzyme does not recognise the 3'overhang and therefore cannot initiate digestion on the opposite strand. At timed intervals, aliquots of the reaction mix are removed, the DNA blunt-ended, and ligated to recircularize the molecules. To sequence the genomic region between the CRH-R2α and CRH-R2γ exons, we used this method (Erase-a Base System Kit, Promega), but we were unsuccessful.

(3) Shotgun sequencing.

The target DNA is fragmented by enzymatic digestion or physical shearing to fragment sizes in the range of 0.5-5 kb. Subtractions with a narrower size range (e.g. 0.8-1.5 kb) are subcloned into either a plasmid or M13 vector. The subclones can then be sequenced from standard primer binding sites in the flanking vector DNA. Shotgun sequencing requires no prior knowledge of the insert sequence, and puts no limitations on the size of the starting DNA to be sequenced. Recently, projects encompassing entire bacterial genomes have been completed using a whole genome shotgun sequencing approach (Fleischmann *et al.*, 1995).

(4) Transposon-based strategies.

An alternative approach that avoids subcloning uses mobile segments of DNA to introduce priming sites into a large segment of DNA (Kasai *et al.*, 1992). A clone containing the target DNA is grown in a bacterial host strain carrying the appropriate element and a gene encoding an enzyme (transposase) that catalyses transposition of the element into the target DNA. In addition to sequences required for transposition, the transposable element carries an antibiotic resistance gene or suppressor tRNA gene that can be used as a selectable marker. Clones containing transposable elements are selected by transforming plasmid DNA isolated from the transposase-containing strain into a non-transposase-containing strain, with co-selection for markers carried by the plasmid vector and the inserted transposable element. An ordered array of insertions can be obtained by using PCR or restriction sites within the element to identify insertions at regular intervals in the target DNA. Priming sites at the ends of each element are then used for sequencing runs extending away from the element into the target DNA.

The genomic clone JP1 was digested with *Sst* I and *Sph* I, fractionated on 1% agarose gel and transferred to Hybond-N filter membrane for Southern blot analysis. The filter was probed with a ^{32}P-labeled cDNA fragment (197 bp) corresponding to the region of human CRH-R2α cDNA as indicated in exon I-II. A JP1-*Sst*I 4.0 kb fragment and a JP1-*Sph* I 3.5 kb fragment were observed. In addition, the JP1-*Sst* I 4.0 kb fragment was digested again with *Nco* I and *Kpn* I. A JP1-*Kpn* I 1.7 kb and *Nco* I 2.5 kb were also observed when a Southern blot of genomic DNA was hybridised with the cDNA

(197 bp) probe from exons I-II. The fragments were subcloned and sequenced. Isolated clones were sequenced using internal primers (Primer walking method) for the gene. Nucleotide sequences were determined by the dideoxy chain termination method with Sequenase version 2.0 T7 DNA polymerase kit (Amersham) and with an automated DNA Sequencer. The nucleic acid sequence of the fragment revealed that, both JP1-*Kpn* I 1.7 kb and *Nco* I 2.5 kb, contained CRH-R2α exon I of 103 bp, CRH-R2α intron I of 137 bp and exon II of 126 bp. The location of CRH-R2α intron I was identified by comparing the genomic DNA with cDNA sequence (Liaw *et al.*, 1996). In order to sequence the genomic region between the CRH-R2α and CRH-R2γ exons, primer walking was employed. The JP1-*Nco* I 2.5 kb fragment contained CRH-R2γ intron I of 608 bp length (ATG translation initiation codon in exon I, which is denoted as nt +1). The JP 1-*Kpn* I 1.7 kb fragment contained CRH-R2γ intron I of 1333 bp length (ATG translation initiation codon in exon I, which is denoted as nt +1), as shown in Figure 7.1.

In order to extend the upstream region of the CRH-R2α gene, we sequenced the JP1-*Sst* I 4.0 kb fragment. The nucleic acid sequence of the fragment revealed that, Jp1-*Sst* I 4.0 kb contained CRH-R2α exon I of 103 bp, CRH-R2α intron I of 137 bp, the first common exon II of 126 bp and the CRH-R2γ intron I of 2281 bp length (ATG translation initiation codon in exon I, which is denoted as nt +1).

To confirm that clone JP1 included the exon I of the CRH-R2γ gene, the synthetic oligonucleotide (S1), 5'-CTGTGCTCAAGCAATCTGCCTAC-3' was used that corresponds to a sequence in exon I. The Southern blot was hybridised with a [γ-^{32}P]dATP end-labeled oligonucleotide and a unique JP-1 *Sph* I 3.2 kb fragment signal was found. This fragment was subcloned and sequenced. The nucleic acid sequence of the fragment revealed that, Jp1-*Sph* I 3.2 kb contained CRH-R2γ exon I of 227 bp, CRH-R2β intron Ib of 205 bp, and CRH-R2γ intron I of 2831 bp length (Figure 7.1). The distance between the CRH-R2γ exon I and the CRH-R2α exon I is 4068 bp. The genomic DNA sequence corresponded exactly to the cDNA of CRH-R2α sequence.

↓ *Sph* I
ATCTATGATCAATTCAATTGTATGCCCCTGAAAAAGGCAACGCATGCAAAATCACAGAGC
AATTATGACTTCAGGCAGGGTTTAAATAGATTATTTCTATTTTATTCATCACAGAGGCAGC

TACATATATTTGAAAAATATTAGGGCATATATTATGAGGCATATTATTTATCCTATCCATA
GATGATATAAAAATCCAAATGGACTATTTATTTATTTTTTATTATTTTTTGTAGGCTGGTC
TTGAATTCCTGTGCTCAAGCAATCTGCCTACCTTGGCTTCCCCAAGTGCTGAGATTA
TGGGTGTGAGCCACTGCACCTGGCCAAGAATCCGAATGGATTCAAAGATACCTTGAA
ATAATTCCTCAATGCAACACACACACATATGCCAGGGTTGGTCAAATGGGAAGAGAG
CCTTGGCCTGAAGACAGGGACCTGGGCTTTCCTCAGCTCTTCTGCCAAGGTATCTGT
CCTTTCTTAGTGACTCACTGGGCTGAAGTCTAGAGCATTCCAATGGGTGCTGGGGATGGG
TTAGTGAACCAGGACCAGCCCTGCCCTTGGGGAGGCATGGGCTGGTAGATGAGACAATGA
ATAAAAAGCAACCTGGATGTGACATAGGCCAGCACCCAGCAGATGGGGTCACCAAGGAG
CTGCATGTCTGAAGGATGAGTGTGGAGCTGTAAGGCCATTTCCAGTGCAGAAATACAAAC
AAGGAACAGAGATAGCGAATGGTTTC*TGGTAGTGACACATTGTTCTGG*AGGGCCTTTCAAA
TGAAGCAGGAGGGTGAGGATGGGGGACGATGATCATGAAGAACCTTCTTTGCAATACCA
AAGATGGTTCCCAGGAATGACATGTTGTTCTGGAGGGGCCTTTCAAATGGAGCAGAAGGG
TGAGGATGAGGGATAGATGATTACGAAGGACCTTCTTTGCAATACCAAAGAGGTGGTCTC
CATCTCAAGGGCAATGGGTTGGCGGCACTCCTA**TGAGGAAGAGTGACAGGATCG**TAAT
TATAAAACATCGTTTTATCTGCCATGAGGAGACAGGGAGCCCATTTGGGCTGCTGCAGAG
ATGAGGATGGCCTGGGAGAAGTAAGGGAGGGAAACCTCTTTTGCTTGTCTGGGAATCCTA
TTGTCCATCTGTAAAACAAGAATGTGGGATTGGATGTCACTGAGAGTCCTTTCTGTCCCCA
CAGCTTAGTCCCATGGTTTGACATGAAGACACCAGGCCAATTCCTCTGTCCTGCAGTGGA
GTAGGATAGGGTGAGGAGGGGGTGCCTGGCCTGGGACCCCTGTCTCCAGTGCTACAGGGC
AAAAGTGCAGCACCTTCAGGCCCCTCTGGAACCTCTGTGCCCTCAGCAACCCCGTCGTAA
TTCCCAGGAAACACCA*CGACTATTCCAGCTGTGCCTG*AATAGACTCCCTCTCTATGTAGTCT
AACCAAGGACAGTCCCTTAGAGAAGCAAAGATGCATCCTGTCCCTTTAAATCCTGTTTTC
CAGCTGACATTTGAGTGGTAGGGGATGAATTAGAGAGAGAATGTGTGTGCATGGTGAAG
ATGCATAACTGTGGCTCTGTGTGTGATTCTGTTGTGCCTACATGCCTGTTTGATGCCATGT
ATTAGGCTTTGCAAGTGTGTCTGAGCTGGCCTGGAACTGTCCACTGCTGCTGCAGCTGAC
ATCAGAGGTGGGCCATGGGATGGGATGCAGGGCACCAGAGGCACCTGCCTTACTCTGCTG
CTCATGGTACACAGGGGTCTTCCAAAGTAC**CTGTCACACTTCCCTGAACCTA**TTTGCCA
ACCTGTCCCCAACAGCTTGGGGACACACAAACTGTTCCAAATACTTATCCCCTAATTCCTG
GCCTCCAGCTGGGATGGGGCTGGCCTGCAGCGCTGGGAACCCATCACTATCCCA*AAGCCT*
*CTAATCTACCTCTGC*TTCTTTAGTTAGCAAAAATGCCCCTGTCTTTGTTTGTTTGTACTTGGA
TTTAGTAAAATTGAGGGAATTTTGGGGCTCTTCCCAATTTCTTCCTCATTTGCCTGTTTGC
AGACACTAAAGTGAGCTGTAAAATCAATTTGTTCCCAAACTGCTACCTTTTCTAGTTTTCC
CTCTGTCACATCTCGAATGACAACTGTGCCTATAATAATAAGATCCATGAAGAAATGGCC
CCACATCCAGGGGACCTCGGGTCTGTGGGTCTGTGGGTATGTGCTCAGACCCCAACTGCT
CATTCAGGATGCAGAACAGCCTTTGACTCTGCCACTGAAATGGTCACTTCCCAGAAGAAT
↓ *Sst* I
CTTTGGTATGTGAGCT**CTCCCAGTTAAGAGCAATCCA**GTTCACCCAGGCCAACCCTCTGT
TTCTGCAGAAATCTGCATGAGATGAGAAGTCCCTGGCCTCTGGTCATTGGGAGCACACCA
CCTCTTGACATAGCTGCATCTATAGGTGCTTTAGGCACCACCTAGGTGGTGGCTGCAGTA
ATGAACAAGGATAATAA*CTTAAGCTTGGCTTAGGATTTC*TCCCTTCCTCTGTGCTGCACCGT
GCTGTGTGAAGGGCTATCCCTTGTATATGAGGGACACTATTGCCCTGTGAATACATGGGC
TGACCTGGCCAGCCACCCTCTGGCCTGTGTGGGACCCTGGGTAACAGGGCTCAGTCTAGG
AGCGGAGGGTAGGACTCTGCCTAGGTCCTCCCCTGCTCCTGGCACTGGATAATATGAAGA
CCAAGGAACCTCTACCTCTGCAGCTGCCCAGAGCTGAGCCTGGGCTCTCACTGTCCATTCA
GGAGGCAGAGGGTAGACCTGGCTGG*CTGCTCAGCATCTTGTTCCTGC*CAGTCAGCAGGCCC
TGGAGAGATTCCAGGGCTCAGC**CCTGGTCTTAGTGTGGTCTCACT**CGCTCTGCCTGGGC
AGTGACAGCACCATTAATATGGAGGCTGGTGAGAGCGGAGCACACAAAAGCAGCCTGCC
TGCTGCTTTGCCTCTCTCTGCCCAGGGCATGGTGCTAGTTCATGGTGGTTTCAGCCTTTCCT
AGCAGCTTAGGTTTATGTGGAGGATGGCAGGGGAACAGGACT*GGTTTCTGAGACTAGGTTC*
*CA*CTCTCCTTGCCCTCTAAAGATAGAAACAAACAACACACAACACACATGTATGTCTTCC
CAGACTCTCCGTCTCATAACTCAGAACCAGAGAGTCTCAGAGCTGAGAGGGGCCTCGGGG
Kpn I ↓
AGATTTCACTGATGGAGAAAGCTCCAGAAGAGAGGGCAAGCATCCTGCCTGGGGTAC**CA**
Sph I ↓
TGGCAGGGAGAGCAGAGTCAGAGGCTGGACCAAAATGCAGCCTGAGTGGAAGCATGCC
TTACCCCAGGCCCTGCTTCCCGAGCCCAATGCCCCACTCAGCTGGCAGACACAGCAGGCC
CGCCCTGGGCACTCTGAGGGTGCAGCCACAGCAGATCACCAAGGAGGCATTCTGGGCCA
GGGTGGGGTGGGGGGCCTGGGCAAGTTCTCTGGGGAGGTTTCCAGCTCCTCCACACCTGC
TGTGGGGCCTGATTCTCCCCGCCCCTGCCCCGCTACTGG*TGTGGAAACCAGGGTCAGGTG*A

AGCCCTGCCCAAGCCTTGAGGAAGAGAGACACATTCTCACTTCTTCCTTTATTCTTTCTGG
TGCCCAAGGCACTAACAACTGGGTGTATAAAGCCTTCCAGATAATTTCAGCCAATTTCTC
CCCTAGACTTACTCCATCTGATTAAATGGCCACCCAGTCACTCAAGCAGGAGACCTGATGT
TATCCCTGCCTCTGGGCTCCTACTTAAAACCAGCCACCCACTCCAGTTCCTCCTCACCAGG
TCTGGCTGCCCCTCCGAAATCTCTCTCCATTCCAGCTTCCACGGTCTGATCCAAGGACTTC
AACCCACTCAGGCCA**CTTGATCTCATTAGACTCATG**GGTGTCTCCCTGCTTCCCTCCAGG

Nco I ↓

CCTCCTCGATACATTTCCTTCCCAGAGGCTATGGAGGCTTTTGGAGGATGCAGATCTGCCC
ATGGGCTCTCCGTCTCTGCCTCCCTCCAGTGTCCTCGCTCAGGGAGGGGAAGCTCAGGCGG
AAAGCTGCCGAACTTTGGGTTGCGGCTGTCCCTCGATTAGCAGAGCTGCGGTGTTCTCCTC
GGGCAGGCGGGCAGGTGGGCGCGCTTTGCTGCCCCCTGCAGCTCGGGGGCCTGCGATCCC
CGCACAGAGCATTCCGTCACCCCAGGCCCACGCTCTCCAGCCCACCGCCCTCCT**CTGGAC**
GCCGCGAGTGGAAGA*GAGCTGCGAACTGAGAAGCC*GTACTTTGGGCAGGGTGGAGGGCC
CGGGGGCTGGAGACTGAGCCCCTCCGAGAGGAGCCGCCCGGCCCCGCCCCCCGGCGCAG
CCATTGGCCGCGGCGGAGCGGCTGTACCCGCAGCTCCGTGCACTCGGCGGCTCCTCTCCG
GGAAGGTCCCCACTTGACAGCTCTGGGCGACCG**GAGGTGGCGCCCAAAGGCTGCC**CGG
GAGATCGGGGCTGGGCTGGCGGGGGCCAGGACCCCGCGCCCTCTCGGCCGCTCACTCTCG
CGTCCACTCCCTCGCAG**TCACGCCGGGCGCGCACTCCCACTCCCTCTCCGCACGCGGC**
TGCGGGACGCGATGGACGCGGCACTGCTCCACAGCCTGCTGGAGGCCAACTGCAGC
CTGGCGCTGGCTGAAGAGCTGCTCTTGGACGGCTGGGGCCACCCCTGGACCCCGAG
GGTAGCGGAAGCGAGCGGTCGAAGCTGCGCGGGCTCCTTCTCTCGCGCGCTGGGTGCGCC
CCTGACCCCGCGCTCCGAGAGCCCCGCGTCCCTGCGGGCCCCTCTGAGCCCTACTGCGCTC
TCTCCCGCCGCCAG**GTCCCTACTCCTACTGCAACACGACCTTGGACCAGATCGGAACG**
TGCTGGCCCCCGCAGCGCTGCCGGAGCCCTCGTGGAGAGGCCGTGCCCCGAGTACTT
CAACGGCGTCAAGTACAACACGACCCGTGAGTGCCCCCGGCCCTCGCTGCCTGCTCT

Figure 7.1 Nucleotide sequence of the genomic region between the CRH-R2α and CRH-R2γ exons.

Antisense and Sense internal primers for the sequences are shown in bold and italic letters (underlined), respectively. CRR-R2α exon I, CRH-R2γ exon I and first common exon II sequences are in bold type, intron sequences are in normal type.

Comparison of the genomic sequence with cDNA obtained by RT-PCR and 5'RACE established the organisation of the 5'end of the CRH-R2α gene. 5'RACE analysis extended the previously known exon I sequence of the human CRH-R2α and established that 5'untranslated region of CRH-R2α mRNA is 52 nt (for details, see Chapter 8). Exon I contains 52 bp of 5'untranslated and 103 bp of translated DNA sequence. Two nucleotides at both ends of the intron completely fulfil the GT-AG rule. This rule claims that dinucleotides GT and AG are invariably present at the 5' and 3' end of introns, respectively. The results suggested that the promoter of the

CRH-R2 is located upstream of the region. The distance between the CRH-R2γ exon I and the CRH-R2α exon I was 4068 bp length.

Between CRH-R2γ I and CRH-R2α I intron (named CRH-R2γ I), sequence data was analysed using ORF (open reading frame) Finder database searches from the National Centre for Biotechnology Information (NCBI). There are numerous ATG codons in the CRH-R2γ I intron sequence leading to seventeen short open reading frames (ORFs). ORF frames range from 105 bp to 546bp. However, using 5'RACE and RT-PCR from cDNA made from some tissues (myometrium, hippocampus) was found no evidence for the presence of additional exons in this region for the human 5'end of CRH-R2 gene.

7.3 Sequence analysis of the genomic region between the CRH-R2β Ib and CRH-R2γ exons

We sequenced the clone (JP1) which includes the exon I of the CRH-R2γ and CRH-R2α genes. The nucleotide sequence of the genomic DNA revealed that JP1-*SphI* 3.2 kb contained the CRH-R2γ exon I of 227 bp and 205 bp of the CRH-R2β intron Ib, but did not contain the CRH-R2β exon I. In order to obtain the genomic sequence between the CRH-R2β and CRH-R2γ exons, we used PCR with Elongase system (Gibco BRL) and human genomic DNA from peripheral blood leucocytes. First, we used a primer to the CRH-R2β exon 5'end (B2:5'-CTGGCATGAGGGGTCCCTCAG-3') and a primer to the 5'-end of the CRH-R2γ specific exon (G3:5'-GTAGGCAGATTGCTTGAGAC-3'), but genomic PCR was unsuccessful. Second, we used a primer to the CRH-R2β exon 3'end (B4:5'-CATGACCCTCACCAACCTCTC-3') and a primer to the 5'end of the CRH-R2γ specific exon (G3:5'-GTAGGCAGATTGCTTGAGAC-3') and generated a 2.6 kb fragment (Figure 7.2). The product was subcloned into the pGEM-T vector and sequenced with SP6 and T7 primers. Sequence analysis of this fragment confirmed the presence of the CRH-R2β and CRH-R2γ exon sequence at the 3'and 5'ends, respectively. The CRH-R2β and CRH-R2γ exons possess good consensus splice junctions flanking an apparent 2575 bp intron.

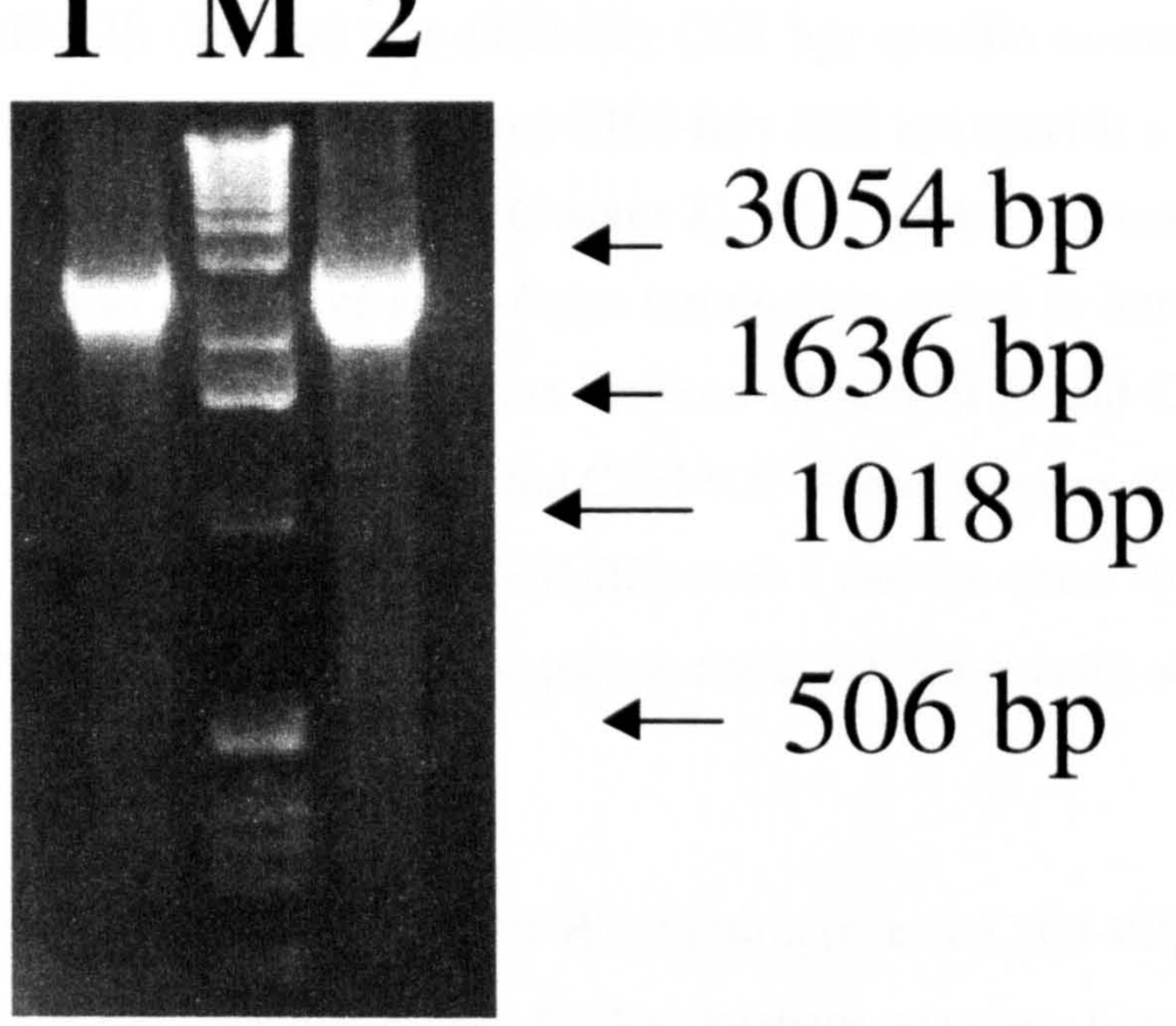


Figure 7.2 PCR was used to amplify the genomic region between the CRH-R2β Ib and CRH-R2γ exons.

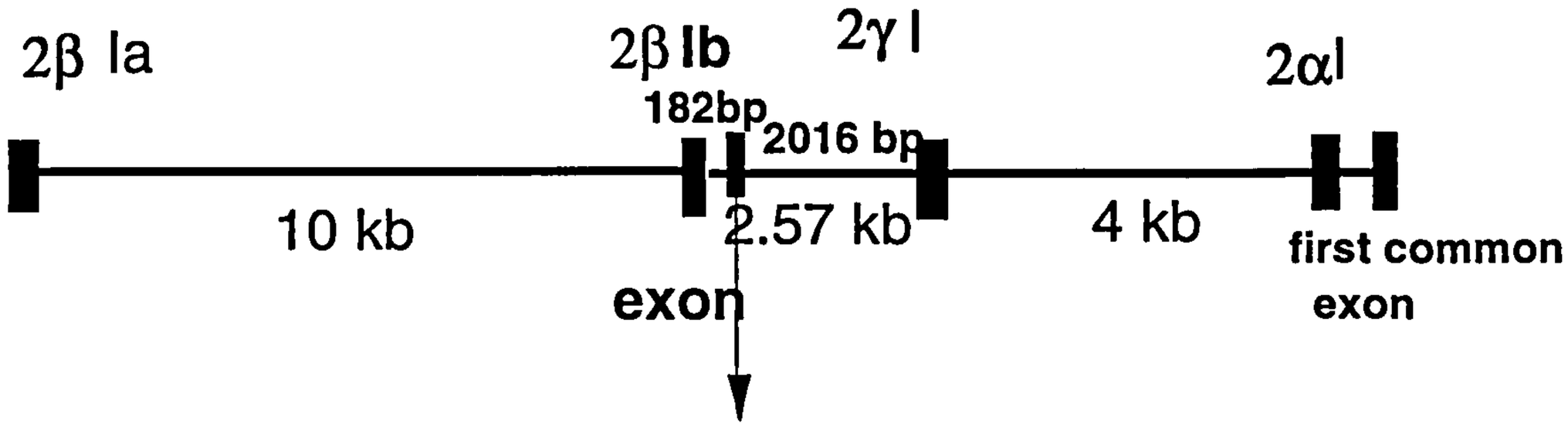
Lane 1 shows (human genomic DNA) PCR amplification of a 2.6 kb fragment, which contained CRH-R2β and 2γ exon sequence at the 3' and 5'ends. M is a 1kb DNA size marker. Lane 2 PCR amplification of a 2.6 kb fragment from the clone JP1.

Comparison of the genomic sequence with cDNA obtained by RT-PCR and 5'RACE established the organisation of the 5'end of the CRH-R2γ gene. The CRH-R2γ cDNA sequence identified was identical to the corresponding genomic DNA sequence.

The resultant products of 5'RACE obtained using CRH-R2γ exon I gene specific primers were cloned into the pGEM-T cloning vector and the clones were sequenced. The nucleic acid sequence of two fragments (A and B) revealed that 488 bp fragment corresponded to the CRH-R2β (280 bp) and CRH-R2γ (208 bp) specific exons, the 690 bp fragment contained CRH-R2β (280 bp) and CRH-R2γ (208 bp) specific exons and an extra 195 bp of intronic sequence (see Chapter 8). This product potentially encodes a truncated protein as a result of an in-frame termination codon in intron I. The extra 195 bp sequence is located at the region between CRH-R2β Ib and CRH-R2γ exon I (Figure 7.3A). The distance between the CRH-R2β Ib and an extra 195 bp sequence is 182 bp. The distance between the CRH-R2γ-exon I and the extra 195 bp sequence is 2016 bp. The extra 195 bp cDNA sequence corresponded exactly to the genomic DNA sequence (Figure 7.3B).

The sequence data between CRH-R2β Ib and CRH-R2γ I intron (named CRH-R2β Ib) was analysed using ORF (open reading frame) Finder database searches from the National Centre for Biotechnology Information (NCBI). There are numerous ATG codons in the CRH-R2β Ib intron sequence leading to seven short open reading frames (ORFs). ORF frames have a range from 102 bp to 366 bp. Using gene specific primers for RT-PCR and 5'RACE, one cDNA fragment of 195 bp was amplified from hippocampus cDNA (see Chapter 8) but not myometrial cDNA in the CRH-R2β Ib intron. The CRH-R2β Ib intron was located in this region for the 5'end of CRH-R2 gene, leading to the possibility of other exons within this region.

A



B

↓

GenomicAGGGCTCAGGCTTGCCACTGATAATGAAGGGCTTGGCTGGGAG
cDNA GGCTCAGGCTTGCCACTGATAATGAAGGGCTTGGCTGGGAG

GenomicGCCACTCCCCATTCCCAGCTCCCACGATGGAAGTCAGCTGGGCT

cDNA GCCACTCCCCATTCCCAGCTCCCACGATGGAAGTCAGCTGGGCT

GenomicGGCATTCAGGAGTCCAAGAAGGGGATGACCCCTCAGCTTTCCAA

cDNA GGCATTCAGGAGTCCAAGAAGGGGATGACCCCTCAGCTTTCCAA

GenomicCTCCAGGGAAACTGGCAGAAGGTGGTAAACATTCTCTGGTTGCC

cDNA CTCCAGGGAAACTGGCAGAAGGTGGTAAACATTCTCTGGTTGCC

Genomic GTGGCTACAGGTGTGACTTCCTGT

↑

cDNA GTGGCTACAGGTGTGACTTCCT

Figure 7.3. (A) The arrangement of the 5'-end of the CRH-R2 cDNA with extra 195 bp. The CRH-R2α, 2β (Ia + Ib), 2γ and extra 195 bp alternative exons are represented as solid boxes.

(B) Nucleotide sequence of human CRH-R2γ 5'RACE-derived cDNA (extra 195bp) and comparison to the corresponding genomic DNA sequence. The positions at which the fragment is inserted are indicated by (↓).

7.4 Analysis of the genomic region between the CRH-R2β 5' and 3'ends exon

Analysis of human genomic DNA by PCR and by sequencing indicates that the CRH-R2β exon lies upstream of the CRH-R2γ and CRH-R2α exons. Our results indicate the genomic order of the alternative exons to be CRH-R2β-2γ-2α. PCR with CRH-R2β exon-5'end and CRH-R2γ exon-3'end primers was used to obtain the genomic DNA fragments, but the PCR was unsuccessful. PCR with CRH-R2β exon-3'end and CRH-R2γ exon-5'end primers was used to amplify the region and generate a 2.6 kb fragment (see Section 7.2). The results of the genomic PCR are most likely due to the large intervening distance between the CRH-R2β at the 5'- and 3'end. The result suggests that the CRH-R2β exon I is interrupted by a large intron in the 5'end of CRH-R2 gene.

The complete sequence of human 7p14-15, including the entire CRH-R2β, 2γ, and 2α gene, was made available recently (GenBank accession number AC004976). Comparison of the genomic sequence with cDNA obtained by RT-PCR and 5'RACE established the organisation of the 5'end of the CRH-R2 gene. The CRH-R2β exon I is divided into exon Ia and exon Ib by a large intron, intron Ia, of 1065bp. The 5'(CAATAC) and 3'(GCAGCC) splice site sequences of CRH-R2β intron Ia include the invariant 5'-GT...AG-3' nt and are in general accord with splice rules. The CRH-R2β cDNA sequence is identical to the corresponding genomic sequence.

The sequence data between CRH-R2β Ia and Ib intron (named: CRH-R2β Ia) was analysed using ORF (open reading frame) Finder database searches from the National Centre for Biotechnology Information (NCBI). There are numerous ATG codons in the CRH-R2β Ia and Ib intron sequence leading to thirty-two short open reading frames (ORFs) ranging from 102 bp to 533bp. However, amplification of this region from cDNA made from some tissues (myometrium, hippocampus) has not indicated the presence of other exons in this region for the 5'end of CRH-R2 gene.

7.5 Alternative exon region of CRH-R2 genomic structure

Alignment of the genomic sequence with the 5'-end cDNA sequence of the CRH-R2 revealed that the 5'end of the gene is composed of 4 exons and 3 introns distributed over 18 kb (Figure 7.3). The entire human CRH-R2β gene consists of at least 13 exons and spans approximately 48 kb. The human CRH-R2γ consists of at least 12 exons and spans approximately 34.5 kb. The human CRH-R2α gene spans approximately 30 kb, with the protein-coding sequence interrupted by 11 introns.

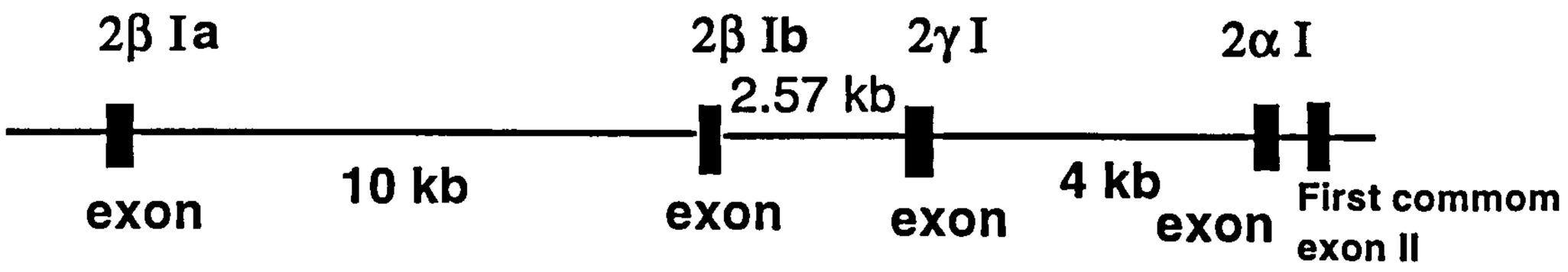


Figure 7.4. The 5'end of the CRH-R2 gene structure.

The 2β exon I is interrupted by a large intron (10659 bp) in the 5'end of CRH-R2 gene. The CRH-R2β Ib and CRH-R2γ exons possess good consensus splice junctions flanking an apparent 2575 bp intron. The distance between the CRH-R2γ exon I and the CRH-R2α exon I is 4068 bp. Accurate estimates of the distance between each exon and the composition of each exon was based on a combination of Southern analysis, DNA sequencing and PCR amplification. The cDNA sequence corresponded exactly to the genomic DNA sequence.

Exon-intron junctions conform to the consensus sequences, with exons flanked by AG and GT at 5' and 3' ends, respectively.

7.6 DNA-binding protein motifs upstream of translation start

We have cloned the 5'-flanking region of the CRH-R2 gene by the PCR based genomic walking method. The nucleotide sequence of 1.3 kb of the 5'flanking region was determined (see Section 6.3.2). A search of the 5'-flanking region sequence against the TRANSFAC (http://transfac.gbf.de/TRANSFAC/index.html.) revealed several potential consensus sequences of putative transcription factor binding sites. The binding site consensus sequences are in shown Figure 7.5.

-1393 AGACCTGTGAGATTTCACAGGAGAGGGAGGTGGGACCTGGAAC

-1350AGCCAGGAAGGCTGCCTGGAGGAAGGGGCTGAGGGTGGGGAGGAAGCACT
GATA-3
-1300 CAAGGACAGAAAAGTGTCCCATGTCTTAGATAGGGAGACCGGAAGGAGCA
CRE-BP SF-1
-1250 CAGAGTGTCTGAGTTACTTAAATCGGGACTGAGTTTTCCAAGGTCTCTCT

-1200 CCACCATGCCTGCATCTCCAGGAGAAAGTTTGGTGCCCCACTAATCACAT
USF
-1150 TCCTCACATGCTTTCCCCTCAGATCCTAACAACAAGCAAGGTGCAACTGT
MYoD
-1100 GGCTTAAGCTTTTAAACCCCCTGCACCAGGCCCACCTGTCAGGAACAGCT

-1050CCTAAGGAGGCTTCTGAGGCTCTGGAAGAACAGACCTCACCTCATGAGCC
SP1
-1000CTGACCCAGCCCCAGGTTGAACCTGTCCCCGCCAAGAACAGCACTGCTCA
CRE
-950 GCCACCTGATATCAGGATCCCTAACCATTTCCTGAAATGGAACCTGGGCC

-900 CCATACAGACAGACACCTGGGCCCCATCCCAAGCTACTTCATTGGAATCT
c-MYb
-850 CTGGGTGAGCCCTGCCATGGCTGTATTTGCTGTCGGTTGGGTTTTTTTTT
Oct-1
-800 TTTTTTTTTGAAAGCATTTTATTTTGATAGAATTTCAAACGTACAGAAAA

HSF2
-750 GTTGGAAGAATGATCCAAAGCATATCCATATACCCCTTTTCCAGATTTAC
C/EBP 1/2GRE
-700 CAACAATTTACATTTTGCCCCATCTGCTCTCTTGTTCTATCTATGTATCT
GATA-1
-650 ATCTGAGTTTATTTCTGAACCATTTGAAAGTATGTTGGAAACATGGTGCC

-600 CCTTTAATCCTAAATGTGTATTTTCCAAGAATGAGAACATTCTCTTACCT

Figure 7.5 Nucleotide sequence of human CRH-R type 2 promoter region with putative binding sites for transcription factors.

The arrows represent the transcription start sites identified by 5'RACE. Putative binding sites for transcription factors are underlined.

SP1 (GC box) = GGGCGG, Oct-1=ATTTGCAT, SF-1=CCAAGGTCA, C/EBP=ATTGCGCAAT, ½GRE=TGTTCT, AP-1=TGAGTCA, AP-2=GCCN$_3$GGC, CRE-BP=TGACGTCA, HSF=CTNGAATNTTCTAGA.

N indicates that any base (A, T, C and G) can be present at that position.

Two Sp1 binding GC-rich motifs are found at -164 and -997. Sp1 sites have been frequently linked to transcriptional control of genes that lack a functional TATA box. There were several DNA motifs characteristic of genes involved in steroidogenesis and reproductive function, including an androgen response element/glucocorticoid response element half site at -665 bp as well as a single steroidogenic factor site at -1204 bp. These include one putative binding site for the enchancer factor AP1

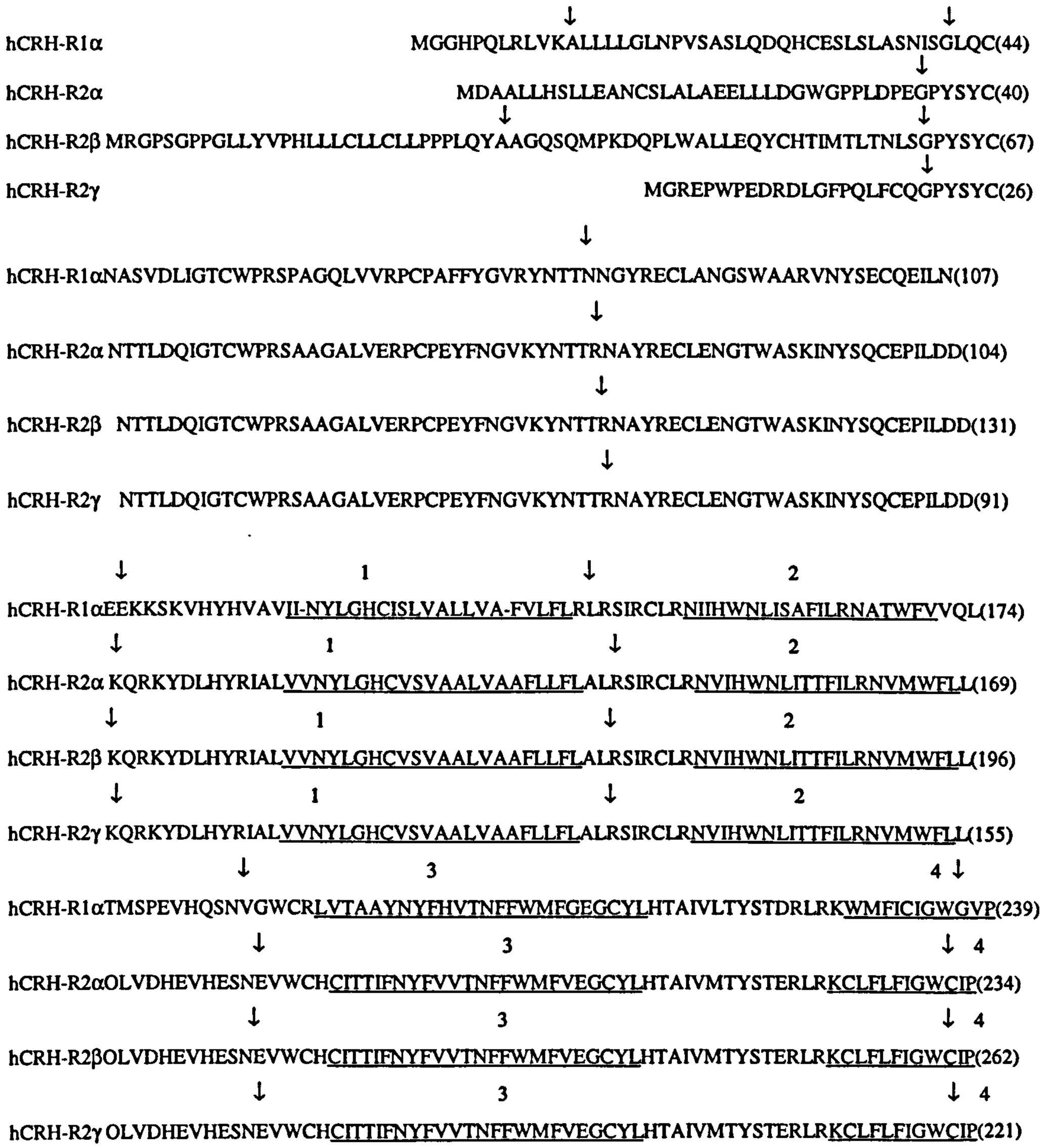
(activator protein-1) at bp -473 and a cAMP-response element CP2 at bp -378 and a binding site for the upstream stimulatory factor USF at bp - 1139. In addition, three putative binding sites for the zinc finger transcription factor GATA are found at -425, -643, and -1285 and a octamer site for the homeobox domain factor Oct-1 at -824bp. The CRH-R2 promoter region up to -1410 bp contains sites for HSF2, deltaEF1, AP-2, and three AP-4. The promoter contains consensus sequences for putative DNA-binding proteins involved in tissue specificity (c/EBP and MyoD) or hormonal regulation (GRE and SF-1). In addition, the promoter region contains consensus motifs corresponding to inducible promoter elements that are known to bind transcription factors induced by exogenous stimuli. These include binding sites for the transcription factors CRE-BP [cAMP-response element (CRE)-binding protein] at bp -1238, CRE at -936.

7.7 **Discussion**

We have determined the structure of the 5'end of the CRH-R2 gene which is organised with the 5'alternative exons arranged as CRH-R2β-2γ-2α. The human CRH-R2α gene spans approximately 30 kb, with the protein-coding sequence interrupted by 11 introns. Genes encoding receptors belonging to the glucagon receptor subfamily are interrupted by several introns. The location of introns in the hCRH-R2β gene was compared with those in reported genes of this subfamily including the human VIP1 (Sreedharan *et al.*, 1995), glucagon (Lok *et al.*, 194), PTH-PTHrP (Kong *et al.*, 1994) and calcitonin (Albrandt *et al.*, 1995) receptors. The N-terminal extracellular domains are separated by four introns in all genes except for that of human PTH-PTHrP-R, which is separated, by five introns. The N-terminal extracellular domains of human CRH-R2α and CRH-R2γ are separated by three introns and CRH-R2β by four introns. Furthermore the arrangement of introns in each domain are quite similar in all these genes. This similarity supports a hypothesis that this subfamily probably emerged early during evolution.

Among CRH-Rs of several species, human CRH-R1's, rat CRH-R1's and human CRH-R2α's gene structures have already been determined (Sakai *et al.*, 1998, Tsai-Morris *et al.*, 1996, Liaw *et al.*, 1996). Rat CRH-R1 and human CRH-R2α genes

contain 13 and 12 exons, respectively. The locations of introns in the coding sequences are identical in human and rat CRH-R1. Human CRH-R2β and CRH-R2γ genes contain 13 and 12 exons, respectively. The N-terminal extracellular domains of human CRH-R1, rat CRH-R1 and human CRH-R2β are divided with four introns, whereas the N-terminal extracellular domain of human CRH-R2α and CRH-R2γ are separated by three introns. However, in the domains toward the C-terminus, the locations of introns are quite similar among the three genes. The similarity of the hCRH-R2 gene to the hCRH-R1gene is illustrated in Figure 7.6.



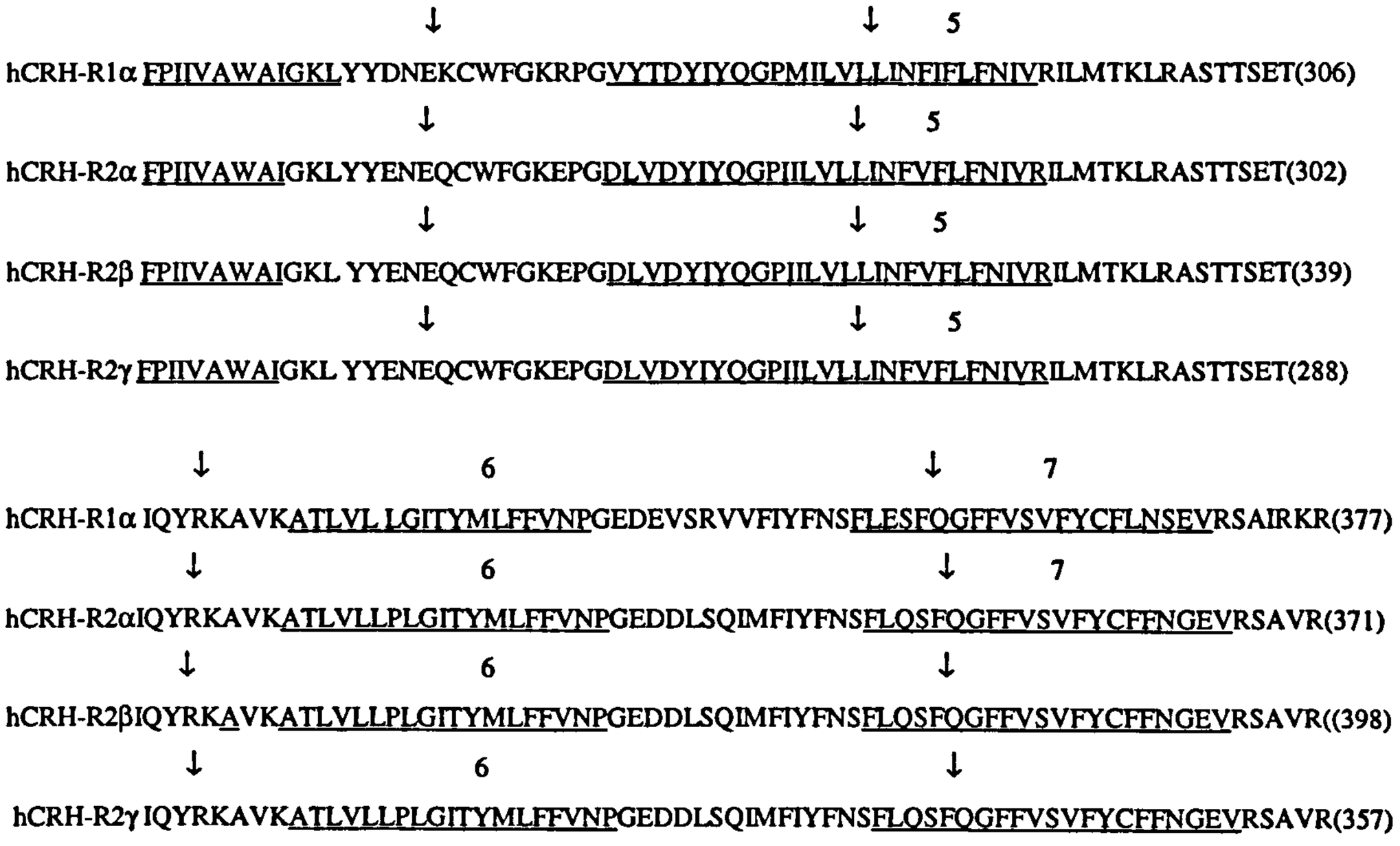


Figure 7.6 Alignment of amino acid sequences of human CRH-R2α (hCRH-R2α) (Accession No. U34587), human CRH-R2β (hCRH-R2β) (Accession No. Y10151), human CRH-R2γ (hCRH-R2γ) (Accession No. AF019381), and human CRH-R1α (hCRH-R1α) (Accession No. L23333) proteins and the location of intron-exon boundaries. Arrows indicate the positions of introns. The putative transmembrane domains are underlined. The numbers in the right-hand column refer to the amino acid number.

Comparison of the genomic sequence with cDNA obtained by 5'RACE and RT-PCR established the organisation of the 5'end of the CRH-R2 gene. The alternative exon region of CRH-R2 gene is composed of 4 exons and 3 introns distributed over 18 kb. The CRH-R2 exon-I is interrupted by a large intron (10659 bp) in the 5'end of the

CRH-R2 gene. Accurate estimates of the distance between each exon and the composition of each exon was based on a combination of Southern analysis, DNA sequencing and PCR amplification. Exon-intron junctions conform with the consensus sequences, with exons flanked by AG and GT at 5' and 3' ends, respectively. Table 7.1 below shows the sequences around the exon-intron junctions.

Table 7.1 Intron-Exon junction Sequences of the Human CRH-R2 5'-end Gene

Exon no	Exon length	5'Splice donor	Intron no	Intron length	3' Acceptor
2βIa	312bp	CAATAC gtaagtg	2βIa	*10659bp	ccctag GCAGCC
2βIb	93 bp	CTCTCAG gtaggt	2βIb	2575bp	ttgtag GCTGGT
2γI	227bp	GCCAAG gtatct	2γI	4068bp	tcgcag TCACGC
2αI	155bp	CCGAGG gtaggc	2αI	137bp	cgccag GTCCCT
II	126bp	CGACCC gtgagt			

Exon and intron sequences are shown in *upper* and *lowercase letters,* respectively. Lengths of the introns and exon were determined by PCR and sequencing, except for one marked with as *asterisk,* which was described by Andrews, *et al* 1998.

The 5'-end cDNA sequence of the CRH-R2 (2β, 2γ, and 2α) corresponded exactly to the genomic DNA sequence. Comparison of the genomic sequence with cDNA obtained by 5'RACE (with CRH-R2α, 2β, and 2γ primers) established the organisation of the 5'end of the CRH-R2 gene. An extra 195 bp sequence can be generated by alternative splicing in the N-terminus of the CRH-R2 gene. The extra 195 bp sequence located in the region between CRH-R2β Ib and CRH-R2γ exon-I. The extra 195 bp cDNA sequence corresponded exactly to the genomic DNA sequence.

The typical eukaryotic gene consists of up to four distinct transcriptional control elements. They are: first, the promoter itself; second, upstream promoter elements located close to it which are required for efficient transcription in any cell type; third, other elements adjacent to the promoter which are interdigitated with the upstream promoter elements (UPEs) and which activate the gene in particular tissues or in response to particular stimuli; and finally, more distant enhancer elements which increase gene activity either in all tissues or in a regulated manner.

The promoter region of most genes contains several elements. The first is a sequence called the TATA box, which located at about 25 to 30 bases upstream from the initial site of transcription (designated as -25 to -30). Some promoters have a second short conserved sequence, instead of or together with a TATA box, which surrounds the start site and is called the Initiator (Inr) element (O'Shea-Greenfield and Smale, 1992). There are two key genetic elements within a core promoter: the TATA box and/or an initiator (Inr) element (Novina and Roy 1996). Core promoter structures can contain both elements (composite: $TATA^{+}$ Inr^{+}), either one element (distinct: containing either $TATA^{+}$ Inr^{-} or $TATA^{-}$ Inr^{+}) or no element (null: $TATA^{-}$ Inr^{-}) (see Chapter 1.7.1). Neither TATA nor other initiator motifs were evident in the 1.3 kb 5'-flanking region (CRH-R2) upstream of the translation starts site.

The absence of TATA and CCAAT motifs is not unique to the CRH-R2: other G protein-coupled receptor genes such as the Rat CRH-R1, the human vasoactive intestinal peptide receptor, the human growth hormone-releasing hormone receptor, and the human secretin receptor (Tsai-Morris, *et al.*, 1996, Sreedharan *et al.*, 1995, Petersenn *et al.*, 1998, Ho *et al.*, 1999) also lack these sequence motifs in their promoters. Promoter regions which are GC-rich and contain Sp1-binding sites but not a TATA box have been shown to be activated by Sp1 recruiting specific co-activators to bind transcription factor II D (TFIID) (Pugh *et al.*, 1990). Pei *et al.*, 1995 have shown that Sp1 binds to the promoter and activates the rat vasoactive intestinal peptide receptor type 1 (VIPR1) gene (Pei *et al.*, 1995). It is possible that a similar transcription mechanism, involving Sp1 rather than TATA binding proteins, operates for the CRH-R2 gene.

The first 1.3 kb of the 5'-flanking sequence contained Sp1 binding sites and lacked a consensus TATA box. All characteristics of gene promoters lacking tissue-specific expression. Moreover, the 5'-flanking region contains consensus-binding sites for several transcription factors that may be functionally important including AP-1, CRE-BP, Ap-2, CP2, GRE and SF-1. From the results of the present study, therefore, we speculate that the strong and housekeeping gene-like activity of the CRH-R2 gene promoter might contribute to the ubiquitous expression of the CRH-R2 gene. Furthermore, protein-binding studies such as footprinting or gel shift analysis will help to verify the involvement of the designated putative sites in the transcriptional control of the CRH-R2 gene.

Some members of the secretin receptor family have multiple untranslated exons in the 5'-untranslated region. The parathyroid hormone (PTH)/PTH-related peptide (PTHrP) receptor gene has three untranslated exons (Manen *et al.*, 1998). The pituitary adenylate cyclase-activating polypeptide (PACAP) receptor gene has four untranslated exons in the 5'-untranslated region (Chatterjee *et al.*, 1997). However, this does not appear to be the case for the CRH-R2. Multiple promoter regions have been described for PTH and glucagon receptor genes. The PTH receptor gene has been shown to be controlled by two promoters, one of which is not GC-rich and is only active in kidney (McCuaig *et al.*, 1995). There is no evidence that the CRH-R2 is controlled by more than one promoter.

Nucleotide sequence analysis of the 5'-flanking region of the CRH-R2 gene indicated the presence of several transcription factors binding sites. There were several DNA motifs characteristic of genes involved in steroidogenesis and reproductive function, including a single steroidogenic factor and an androgen response element/glucocorticoid response element. The 5'-flanking region of the CRH-R2 gene showed a putative binding site for the steroidogenic factor 1 (SF-1) located at from -1213 to 1204 bp (CCAAGGTCT) that corresponded to the reported SF-1 binding elements (PyCAAGGPyCPu) (Lynch *et al.*, 1993). SF-1 is an orphan nuclear receptor, which regulates many genes directly involved in steroidogenesis or other target genes from the hypothalamo-pituitary-gonadal axis (Zhang and Mellon 1996, Sugawara *et al.*, 1997). SF-1 is also implicated in the development of the adrenal gland and gonads (Sadovsky *et al.*, 1995). These reports suggest an important role for

SF-1 in the regulation of steroidogenesis at a number of levels. Further analysis will be conducted to identify the steroidogenic factor 1 within the CRH-R2 gene.

Glucocorticoids are the final effectors of the hypothalamo-pituitary-adrenal (HPA) axis and participate in the control of whole body homeostasis and the organism's adaptive response to stress. These hormones exert their effects through their nuclear receptors. Nuclear receptor molecules bind to short stretches of DNA and consensus sites for the receptor binding are based on certain arrangements of these half-sites. There is an androgen response element/glucocorticoid response element (GRE) half site located at -665 bp in the 5'-flanking region of the CRH-R2 gene. Glucocorticoids play a key regulatory role in the biosynthesis and release of CRH (Otth 1992). Further studies are required to substantiate the functionality of the half-GRE present in the 5'-flanking region of the CRH-2 gene.

A number of putative binding sites for tissue specific factors (see Section 1.7.3.1) are located within the genomic 5'-flanking sequence, for instance, a site for the transcription factor GATA-1 (NF-E1), which is expressed at very high levels in erythroid cells. GATA-1 was originally identified on the basis of its binding to the promoter or enhancer regions of chicken and mammalian globin gene, and was shown to be essential for their erythroid-specific pattern of gene expression, deletion or mutation of GATA-1 binding sites resulting in the abolition of such expression. This factor plays a critical role in the regulation of a number of different erythroid-specific genes and thereby in erythroid cell differentiation. One binding site for tissue specific factor MyoD activates muscle-specific genes. MyoD was not identified on the basis of studying genes of this type. It was isolated on the basis of the finding that its artificial expression within undifferentiated fibroblast-like cells was sufficient to transform them into muscle precursor cells. Hence this factor is not only capable of switching on muscle-specific genes but also, by doing so, actually plays a central role in the production of the differentiated cell type itself.

The CRH-R2 belongs to the family II of G_S -protein-coupled cell surface receptors that activate adenylate cyclase, resulting in a creased cAMP levels and activation of protein kinase A. Nucleotide sequence analysis of the 5'-flanking region of the CRH-R2 gene indicates the presence of several cAMP response element. These include

binding site for the CREB [cAMP-response element (CRE)-binding protein] at bp -1238. A 5'-GCCCAAGGC-3' sequence was found at position -106, corresponding to consensus sequence that binds the transcription factor AP-2 (Imagawa *et al.*, 1987). The metallothionein IIA gene binds a transcription factor referred to as activator protein 2 (AP-2). AP-2 acts as a basal level enhancer (Mitchell *et al.*,1987), but it also stimulates transcription in response to both protein kinase C and cAMP (Imagawa *et al.*,1987). AP-2 sites have been identified in a number of genes, allowing derivation of the relatively variable GC-rich consensus 5'-GCCN$_3$GGC-3' (Williams and Tjian, 1991). The action of AP-2 is regulated developmentally in a cell-specific fashion (Williams *et al.*, 1988). An important step in the study of transcription factors has been the discovery that many constitute the terminal targets of specific signal transduction pathways. The two major signal transduction systems utilise cAMP and diacylglycerol (DAG) as secondary messengers. Each of these pathways is also characterised by the mobilisation of a specific protein kinase (protein kinase A and protein kinase C, respectively) and an ultimate DNA control element [cAMP-responsive element (CRE) and TPA-responsive element (TRE) respectively]. Although these two pathways were initially characterised as distinct systems, accumulating evidence points to extensive 'cross-talk' between them (see Chapter 1.7.4). Intracellular levels of cAMP are primarily regulated by adenylate cyclase. This enzyme is in turn modulated by various extracellular stimuli mediated by receptors and their interaction with G proteins (McKnight *et al.*, 1988). The binding of a specific ligand to a receptor results in the activation of the cAMP pathway. CREB belongs to a family of nuclear proteins CREB-ATF that binds to an 8-bp consensus sequence, TGACGTCA (Deutsch *et al.*, 1988). Activation of signalling pathways through protein kinase A modifies the phosphorylation state of CREB and increases its ability to activate transcription.

A related but distinct DNA-binding protein family known as AP-1 or TRE [12-O-tetradecanoyl phorbol-13-acetate (TPA)-response element] also binds to a unique *cis*-acting DNA element. At position –471, the sequence 5'-T<u>C</u>AGTCA-3' was present, which differs by only a single base (underlined) from the reported AP-1 site (5'-TGAGTCA-3'). Activation of the intracellular protein kinase C pathway is thought to alter the phosphorylation state and ability of AP-1 to activate transcription. There is

recent evidence that both CREB and AP-1 may interact on similar types of DNA sequences, allowing for 'cross-talk' between these two factor (Hai *et al.*, 1991).

Examination of the CRH-R2 promoter revealed the presence of several cAMP response elements and TPA response elements (CREB, AP-1 and AP-2). Further experiments would be necessary to demonstrate the involvement of one or several of these elements or other ones in the cAMP and DAG regulation.

Chapter 8 Determination of transcriptional start sites using 5'RACE

8.1 Introduction

5'Rapid amplification of cDNA ends (5'RACE) is a procedure for amplification of nucleic acid sequences from a messenger RNA template between a defined internal site and unknown sequence of the 5'-end of the mRNA. Other researcher have also described the methodology of amplification with single-sided specificity as "one-sided" PCR (Ohara *et al.*, 1989) or "anchored" PCR (Loh *et al.*, 1989). In general, PCR amplification of relatively few target molecules in a complex mixture requires two sequence-specific primers that flank the region of sequence to be amplified. However, to amplify and characterise regions of unknown sequence, this requirement imposes a severe limitation. 5'RACE methodology offers a possible solution to this problem.

5'RACE, or "anchored" PCR, is a technique that facilitates the isolation and characterisation of 5'region from low-copy messages (Frohman 1988 and Dumas 1991). Although the precise protocol varies among different users, the general strategy remains consistent. The first strand cDNA synthesis reaction is primed using an oligonucleotide complementary to sequences in the gene. Tailing-mediated anchor PCR is a technique in which the terminal deoxynucleotidyl transferase is used to add homopolymeric tails to the 3'-ends of the cDNA. A typical PCR reaction follows, using an anchor primer complementary to the newly added tail and another primer to a known sequence within the gene.

Ligation-mediated anchor PCR is another technique that uses single-stranded ligation of cDNA (SLIC) with T4 RNA ligase to covalently join an oligonucleotide anchor site adjacent to the 3'-end of the cDNA. The single-stranded –cDNA (ss-cDNA) can then be amplified by PCR.

5'RACE requires a very small amount of primary material, and provides a rapid feedback on the generation of the desired product. Information regarding alternative promoters and splicing sequences can be obtained and a judicious choice of primers (e.g., within an alternately spliced exon) can be used to amplify a subpopulation of

cDNAs from a gene for which the transcription pattern is complex. Furthermore, differentially spliced or initiated transcripts can be separated by electrophoresis and cloned separately, and essentially unlimited numbers of independent clones can be generated to examine rare events (Loh *et al.*, 1989). For 5'end amplification, the ability of reverse transcriptase to extend cDNAs all the way to the 5'end of a message is greatly increased, because a primer extension library is created instead of one for a general purpose. In addition, random hexamer-primed cDNA has been adapted to 5'RACE for amplification and cloning of multiple genes from a single first strand synthesis reaction (Harvey *et al.*, 1991). Products of 5'RACE reactions can be used for the preparation of probes. The RACE procedures may be utilised in conjunction with exon trapping methods (Buckler *et al.*, 1991) to enable amplification and subsequent characterisation of unknown coding sequences.

In this chapter, the 5'RACE method was employed using specific PCR primers for each CRH-R2 subtype, to identify the transcriptional start sites of the CRH-R2 gene. A schematic representation of the 5'RACE strategy is shown in Figure 8.1.

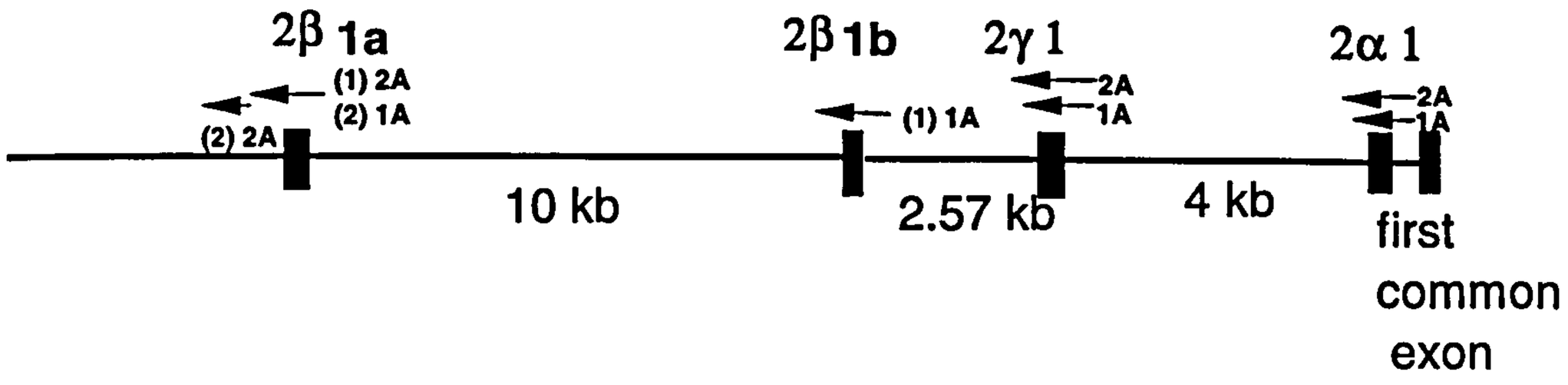


Figure 8.1 A schematic representation of 5'RACE strategy

The CRH-R2α, 2β, and 2γ alternative exons are represented by solid boxes.

The figure illustrates diagrammatically the approximate locations of oligonucleotides used as primers for 5'RACE. CRH-R2α: 1A is first round PCR specific primer. 2A is second round PCR specific primer. The primers are located in CRH-R2α exon-I. CRH-R2γ: 1A is first round PCR specific primer. 2A is second round PCR specific primer. The primers are located in CRH-R2γ

exon-I. CRH-R2β 1b: (1)1A is first round PCR specific primer. (1)2A is second round PCR specific primer. CRH-R2β 1a: (2)1A is first round PCR specific primer. (2)2A is second round PCR specific primer. The primers are located in CRH-R2β exon-Ib. Approximate locations of oligonucleotides are shown by *arrows*.

8.2 Identification of the transcriptional start sites

8.2.1 5'RACE with CRH-R2α exon-I gene specific primers

The 5'end of the human CRH-R2 cDNA was obtained by 5'RACE using the commercially prepared Marathon Ready cDNA system (Clontech) according to the manufacturer's recommendations. An uncloned library of the adoptor-ligated cDNA prepared from human hippocampus was used to amplify the 5'-end of the CRH-R cDNA. First round PCR was performed with a CRH-R2α cDNA specific primer (located in exon-I, + 55/80) and the anchor primer AP1 supplied by clontech. A nested PCR was then performed with the clontech primer AP2 and a second cDNA specific primer (located in exon-I, + 35/61). The products of the second PCR reaction were purified from a 1.0% agarose gel using "QIAquick Gel Extraction kit (Qiagen)" ligated using T4 DNA ligase kit into the plasmid pGEM-T and the clones sequenced.

Next we performed nested PCR of the cDNA using a second nested primer complementary to nucleotide + 36 to + 61 of the CRH-R2α cDNA sequence and an anchor primer. The amplification resulted in two DNA fragments of apparently 309 bp and 536 bp (Figure 8.2).

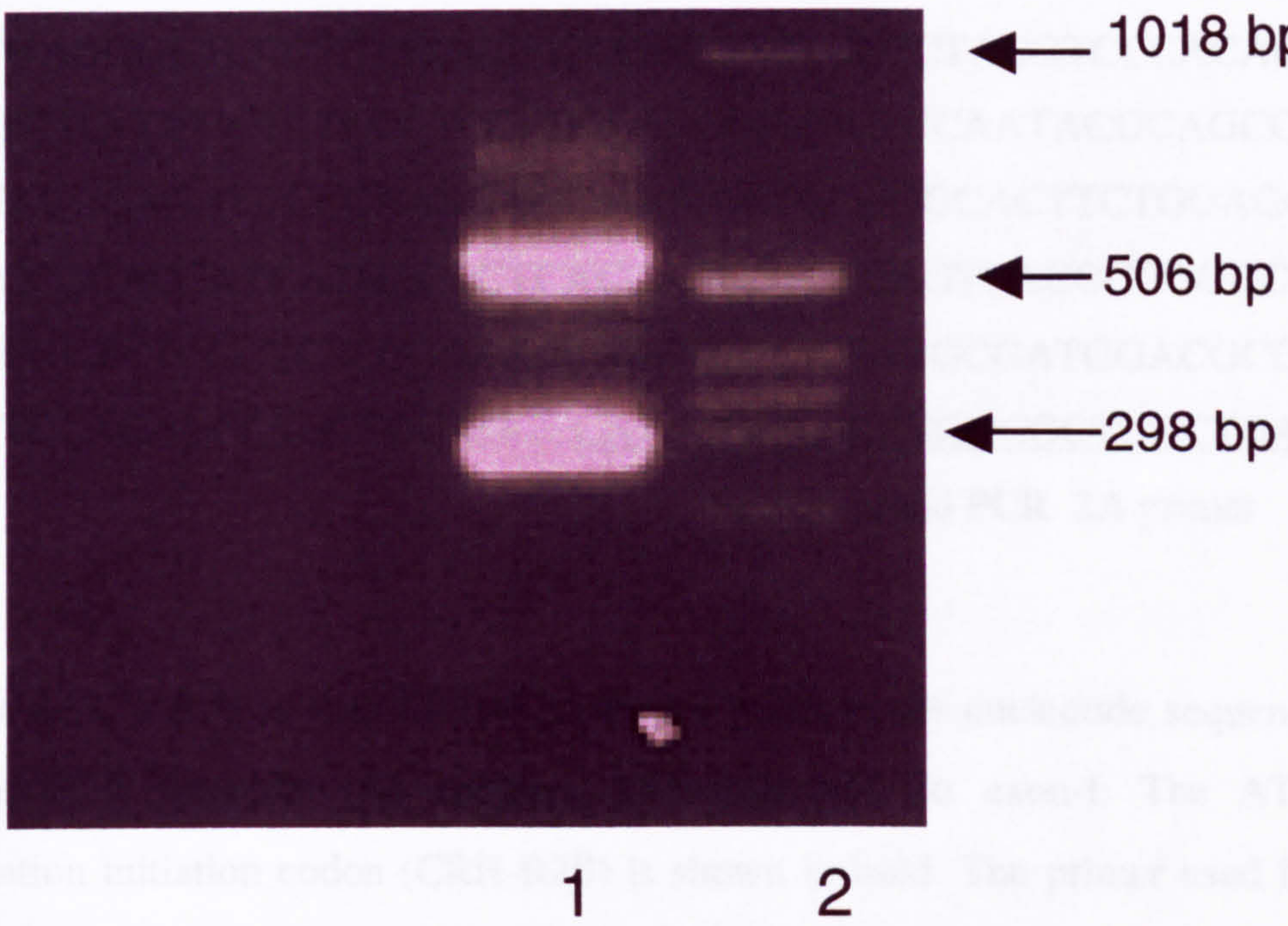


Figure 8.2 Determination of transcriptional start sites using 5'RACE with CRH-R2α exon I primer

5'RACE was used to identify the transcriptional start sites of the CRH-R2 gene. Two cDNA fragments were generated, which have the following sizes: 309 (A) and 536 (B) bp. Human hippocampus cDNA (Lane 1). Lane 2 are DNA size markers.

The nucleotide sequence has been submitted to GenBank under accesion numer AF 361106.

The nucleic acid sequence of one fragment revealed that the 309 bp fragment (A) corresponded to the CRH-R2β (196 bp) and CRH-R2α (113 bp) exon-I (Figure 8.3a).

GCCACTACT
GGCATGAGGGGTCCCTCAGGGCCCCCAGGCCTCCTCTACGTCCCACACCT
CCTCCTCTGCCTGCTCTGCCTCCTCCCACCGCCGCTCCAATACGCAGCCG
GGCAGAGCCAGATGCCCAAAGACCAGCCCCTGTGGGCACTTCTGGAGCAG
TACTGCCACACCATCATGACCCTCACCAACCTCTCAGTCACGCCGGGCGC
GCACTCCCACTCCCTCTCCGCACGCGGCTGCGGGACGCGATGGACGCGGC
ACTGCTCCACAGCCTGCTGGAGGCCAACTGCAGCCTGGCGCTGGCTGAAG
Second round PCR 2A primer

Figure 8.3a. 5'RACE with CRH-R2α exon-I primers, the nucleotide sequence of the PCR fragment (A) contains CRH-R2β and 2α exon-I. The ATG translation initiation codon (CRH-R2β) is shown in bold. The primer used for 5'RACE (second round PCR 2A primer) is underlined.

The sequence of another fragment (B) contained CRH-R2β (196 bp), CRH-R2γ (227 bp) and CRH-R2α (113 bp) exon-I (Figure 8.3b). DNA sequencing of both fragments revealed that the 5'end of the CRH-R2 transcript is located at 12 bp upstream of the CRH-R2β of the ATG translation start site.

GCCACTACTGGCATGAGGGGTCCCTCAGGGCCCCCA
GGCCTCCTCTACGTCCCACACCTCCTCCTCTGCCTGCTCTGCCTCCTCCC
ACCGCCGCTCCAATACGCAGCCGGGCAGAGCCAGATGCCCAAAGACCAG
CCCCTGTGGGCACTTCTGGAGCAGTACTGCCACACCATCATGACCCTCAC
CAACCTCTCAGGCTGGTCTTGAATTCCTGTGCTCAAGCAATCTGCCTACCT
TGGCTTCCCCAAGTGCTGAGATTATGGGTGTGAGCCACTGCACCTGGCCA
AGAATCCGAATGGATTCAAAGATACCTTGAAATAATTCCTCAATGCAACA
CACACACATATGCCAGGGTTGGTCAAATGGGAAGAGAGCCTTGGCCTGAA
GACAGGGACCTGGGCTTTCCTCAGCTCTTCTGCCAAGTCACGCCGGGCGC

GCACTCCCACTCCCTCTCCGCACGCGGCTGCGGGACGCGATGGACGCGGC

ACTGCTCCACAGCCTGCTGGAGGC<u>CAACTGCAGCCTGGCGCTGGCTGAAG</u>

Second round PCR 2A primer

Figure 8.3b 5'RACE with CRH-R2α exon-I primers, the nucleotide sequence of the PCR fragment (B) containing CRH-R2β, 2γ, and 2α exon-I. The ATG translation initiation codon (CRH-R2β) is shown in bold. The primer used for 5'-RACE (second round PCR 2A primer) is underlined.

The nucleotide sequence has been submitted to GenBank with accession AF 361107.

These results confirmed the genomic order of the alternative exons to be CRH-R2β-2γ-2α. The mature receptor subtype mRNA appears to be generated via 5' exon differential splicing.

8.2.2 5'RACE with CRH-R2γ exon-I gene specific primers

5'RACE was performed using the Marathon cDNA amplification Kit (Clontech) with human hippocampus cDNA. The cDNA was initially amplified using the kit supplied anchor primer and the CRH-R2γ specific primer. The reaction products were then further amplified using a supplied nested anchor primer and a nested CRH-R2γ specific primer. Specific primers for the CRH-R2γ 5'RACE have been described in Section 2.18. Two bands of 495 and 690 bp in length were demonstrated by gel electrophoretic analysis as shown in Figure. 8.4. The bands were eluted from the gel and subcloned into the pGEM-T vector. For each cDNA fragment, the nucleotide sequence of six independent clones was determined, which was the same for all clones.

The nucleic acid sequence of two fragments (A and B) revealed that 488 bp fragment corresponded to the CRH-R2β (280 bp) and CRH-R2γ (208 bp) specific exon-I (Figure 8.4), another fragment corresponded to the CRH-R2β (280 bp), CRH-R2γ (208 bp) specific exons and an extra 195 bp intronic sequence.

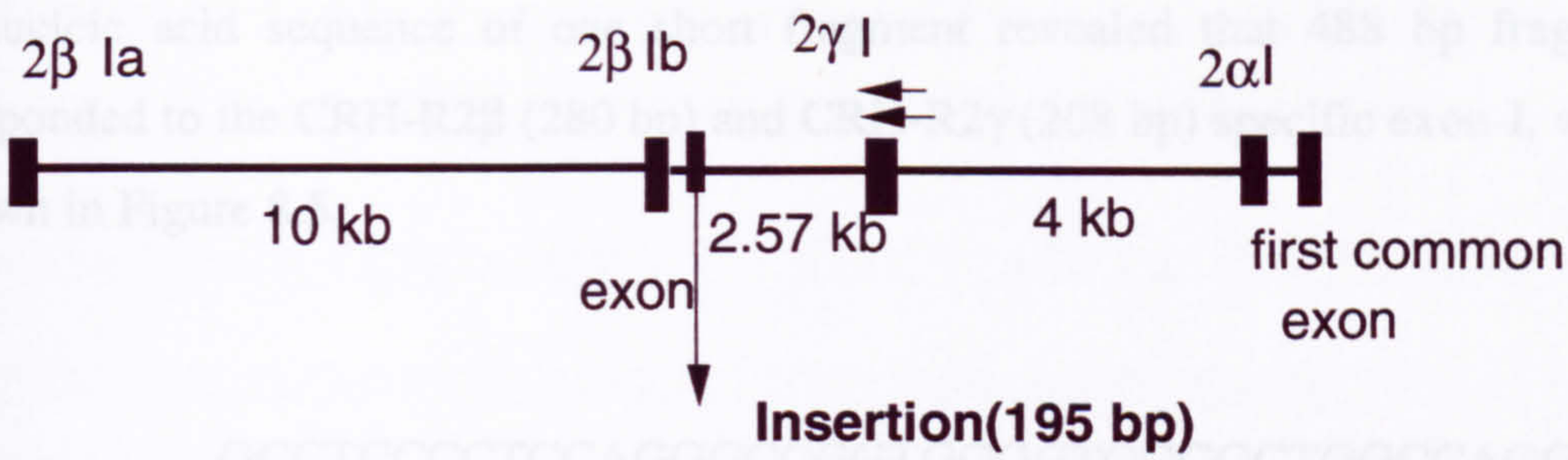


Figure 8.4 The arrangement of the 5'-end of the CRH-R2 cDNA (5'RACE with CRH-R2γ primers). 5'RACE was used to identify the transcriptional start sites of the CRH-R2 gene. The CRH-R2α, 2β, and 2γ alternative exons are represented by solid boxes. (A)RT-PCR product containing CRH-R2β, 2γ and 2α exon-I. (B) RT-PCR product containing CRH-R2β, 2γ, 2α exon-I and insertion (195bp). (C) Lane 2, 3, and 4 are different concentrations of human hippocampus cDNA. Lane 1 and 5 are DNA size markers.

The nucleic acid sequence of one short fragment revealed that 488 bp fragment corresponded to the CRH-R2β (280 bp) and CRH-R2γ (208 bp) specific exon-I, which is shown in Figure 8.5.

GCCTCCCCTCCAGGGCCGGTGGGGTGGGGCTGGCCAGGG
TGTGACCACCGTGCTGGGCAGCAGGCTCCAGTCCCTAACCCCCAGCCACT
ACTGGCATGAGGGGTCCCTCAGGGCCCCCAGGCCTCCTCTACGTCCCACA
CCTCCTCCTCTGCCTGCTCTGCCTCCTCCCACCGCCGCTCCAATACGCAG
CCGGGCAGAGCCAGATGCCCAAAGACCAGCCCCTGTGGGCACTTCTGGAG
CAGTACTGCCACACCATCATGACCCTCACCAACCTCTCAGGCTGGTCTTG
AATTCCTGTGCTCAAGCAATCTGCCTACCTTGGCTTCCCCAAGTGCTGAG
ATTATGGGTGTGAGCCACTGCACCTGGCCAAGAATCCGAATGGATTCAAA
GATACCTTGAAATAATTCCTCAATGCAACACACACACATATGCCAGGGTT
GGTCAAATGGGAAGAGAGCCTTGGCCTGAAGACAGGGACCTGGGCTTT

Second round PCR 2A primer

Figure 8.5 5'RACE with CRH-R2γ exon-I gene specific primers
The nucleotide sequence of the PCR fragment contains CRH-R2β, and CRH-R2γ exon-I. The ATG translation initiation codon (CRH-R2β) is shown in bold. The primers used for 5'-RACE (second round PCR 2A primer) is underlined.

The sequence of another long fragment contained CRH-R2β (280 bp), CRH-R2γ (208 bp) specific exons and an extra 195 bp (55 amino acids) of intronic sequence. The cDNA inserted between the CRH-R2β and CRH-R2γ exons, potentially, encodes a truncated protein as a result of an in-frame termination codon in intron I as illustrated in Figure 8.6.

GCCTCCCCTCCAGGGCCGGGTGGGGTGGGGCTGGC
CAGGGTGTGACCACCGTGCTGGGCAGCAGGCTCCAGTCCCTAACCCCCAG
CCACTACTGGCATGAGGGGTCCCTCAGGGCCCCCAGGCCTCCTCTACGTC

CCACACCTCCTCCTCTGCCTGCTCTGCCTCCTCCCACCGCCGCTCCAATA

CGCAGCCGGGCAGAGCCAGATGCCCAAAGACCAGCCCCTGTGGGCACTTC

↓

TGGAGCAGTACTGCCACACCATCATGACCCTCACCAACCTCTCAGGCTCA

GGCTTGCCACTGATAATGAAGGGCTTGGCTGGGAGGCCACTCCCCATTCC

CAGCTCCCACGATGGAAGTCAGCTGGGCTGGCATTCAGGAGTCCAAGAAG

GGGATGACCCCTCAGCTTTCCAACTCCAGGGAAACTGGCAGAAGGTGGTA

↓

AACATTCTCTGGTTGCCGTGGCTACAGGTGTGACTTCCTGGCTGGTCTTG

AATTCCTGTGCTCAAGCAATCTGCCTACCTTGGCTTCCCCAAGTGCTGAG

ATTATGGGTGTGAGCCACTGCACCTGGCCAAGAATCCGAATGGATTCAAA

GATACCTTGAAATAATTCCTCAATGCAACACACACACACATATGCCAGGGTT

GGTCAAATGGGAAGAGAGCCTTGGCCTGAAGACAGGGACCTGGGCTTT

Second round PCR 2A primer

Figure 8.6 5'RACE with CRH-R2γ exon-I primers, the nucleotide sequence of the PCR fragment contains CRH-R2β, 2γ exon-I and an extra 195 bp intronic sequence. The ATG translation initiation codon (CRH-R2β) is shown in bold. The primers used for 5'-RACE (2A) is underlined. The positions at which the fragment is inserted are indicated by (↓).
The nucleotide sequence data have been deposited in GenBank under No AF 361108.

DNA sequence of both fragments revealed that the 5'end of the CRH-R2 transcript is located at 97 bp upstream of the CRH-R2β of the ATG translation initiation site.

8.2.3 5'RACE with CRH-R2β exon-I gene specific primers

To study the intact untranslated cDNA 5'end of the human CRH-R2 receptor, a Primer 1A (located in CRH-R2β exon-I) and a nested primer 2A (located in CRH-R2β exon-I) were used for the first and second amplifications, respectively, plus an anchor primer AP1 supplied by Clontech. For approximate primer locations, see

Figure. 8.1. The resultant products were cloned into the pGEM-T cloning vector and the clones were sequenced. DNA sequencing of PCR products revealed that the 5'-end of the CRH-R2 transcript is a cytosine residue located at 201 bp upstream of the ATG translation start site.

In order to confirm the 5'end cDNA region, we designed a new primer (located at 121-146 bp upstream of the ATG translation initiation site) based on the sequences of the above clones. Of those fragments amplified with the new primer, the products were characterised by cloning, sequencing, and analysis, which are shown in Figure 8.7.

GCTGGGTAGC

AGCAGCTGTCTGGGACCAGCTATAAATAACCTGGAGGCTGTGGGCTGTGG

↑

GGCGGGGTCACCACTGCCCTGTGGAAGAAGTGGGTGGAGAATACCTCCTC

(2) 2nd: 2A primer

ATGCCCAAGGCCTTACCTTGGCCTCCCCTCCAGGGCCGGGTGGGGTGGGG

CTGGCCAGGGTGTGACCACCGTGCTGGGCAGCAGGCTCCAGTCCCTAACC

CCCAGCCACTACTGGCATGAGGGGTCCCTCAGGGCCCCCAGGCCTCCTCT

ACGTCCCACACCTCCTCCTCTGCCTGCTCTGCCTCCTCCCACCGCCGCTC

(2)1st:1A.(1)2nd:2A,prime

CAATACGCAGCCGGGCAGAGCCAGATGCCCAAAGACCAGCCCCTGTGGG

CACTTCTGGAGCAGTACTGCCACACCATCATGACCCTCACCAACCTCTCA

(1) 1st:1A primer

Figure 8.7 5'RACE with CRH-R2β exon-I gene specific primers

The nucleotide sequence of the PCR fragment contains CRH-R2β exon-I. The ATG translation initiation codon (CRH-R2β) is shown in bold. The primers used for 5'-RACE (1A and 2A) are underlined. The arrow represents the transcription start site.

Apart from a faint band supporting the finding that transcription initiation occurs at C (- 201), an additional strong band was visible, which corresponds to G (- 226). Analysis of the 5'RACE products is shown in an ethidium bromide-stained gel in Figure 8.8.

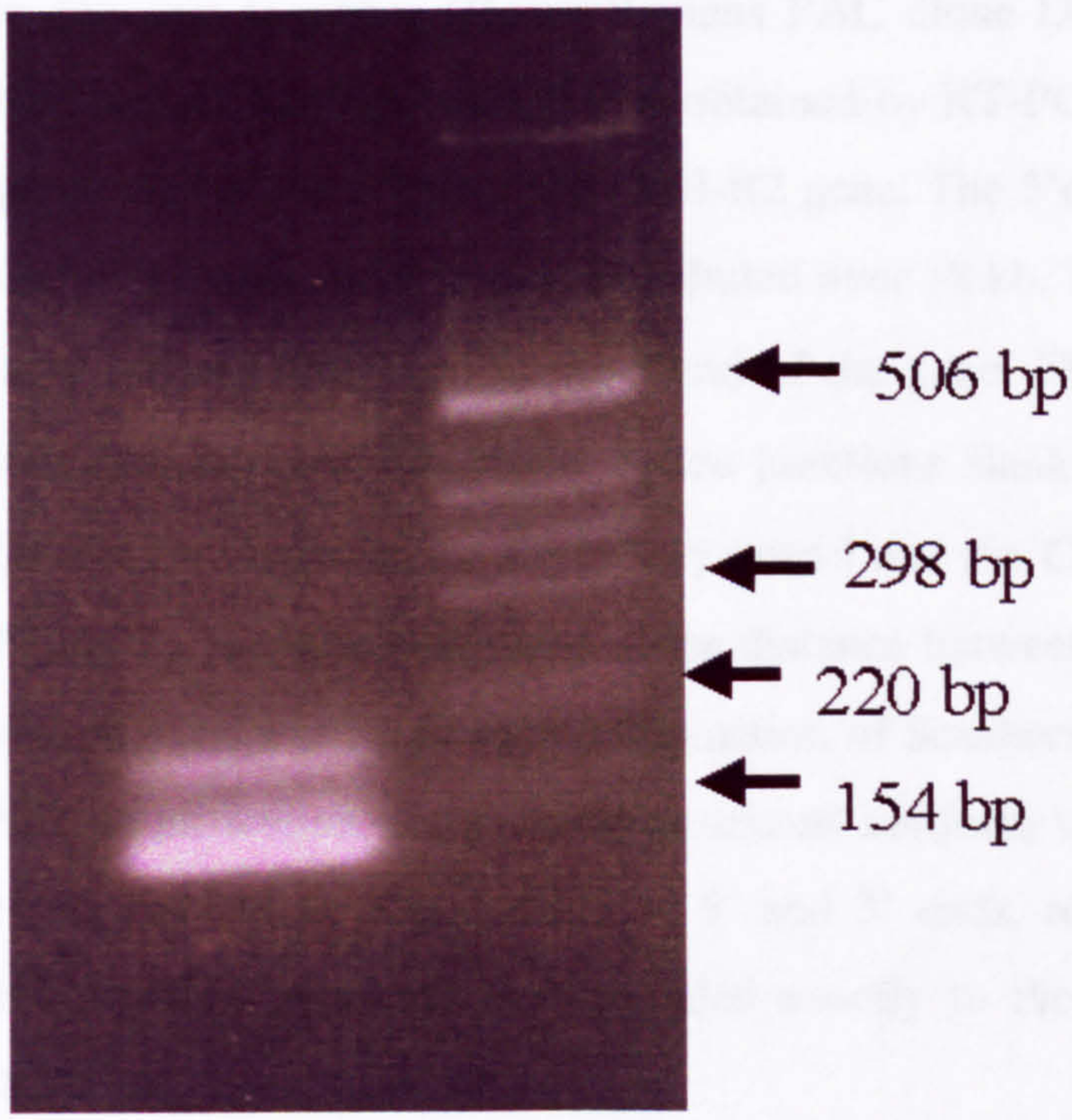


Figure 8.8 Determination of transcriptional start sites using 5'RACE with CRH-R2β exon I primer

5'RACE was used to identify the transcriptional start sites of the CRH-R2 gene. Two cDNA fragments were generated, which have the following sizes: 103 and 128 bp. cDNA prepared from human hippocampus cDNA (Lane 1). Lane 2 are DNA size markers.

The nucleotide sequence of the fragments revealed that the 5'end of the CRH-R2 transcript are located at 201 and 226 bp upstream of the CRH-R2β of the ATG translation start site.

8.3 Alignment 5'RACE-derived cDNA and genomic DNA sequences

Comparison of the genomic sequence (Homo Sapiens PAC clone DJ1143H19 from 7p14-p15, GeneBank No: AC004976) with cDNA obtained by RT-PCR and 5'RACE established the organisation of the 5'end of the CRH-R2 gene. The 5'end of the CRH-R2 gene is composed of 4 exons and 3 introns distributed over 18 kb. The 2β exon-I is interrupted by a large intron (10659 bp) in the 5'end of the gene. The CRH-R2β Ib and CRH-R2γ exons possess good consensus splice junctions flanking an apparent 2575 bp intron. The distance between the CRH-R2γ exon I and the CRH-R2α exon I is 4068 bp (see Chapter 7). Accurate estimates of the distance between each exon and the composition of each exon was based on a combination of Southern analysis, DNA sequencing and PCR amplification. Exon-intron junctions conform to the consensus sequences, with exons flanked by AG and GT at 5' and 3' ends, respectively. The 5'end of the CRH-R2 cDNA sequence corresponded exactly to the genomic DNA sequence (Figure 8.9).

```
Genomic   GCAGGCTGGATGGAGAGCATAGCCCCTTGGCTGGCTGGGTAGCAGCAGCT ↑
5'-RACE                                          GCTGGGTAGCAGCAGCT
Genomic   GTCTGGGACCAGCTATAAATAACCTGGAGGCTGTGGGCTGTGGGGCGGGG
5'-RACE   GTCTGGGACCAGCTATAAATAACCTGGAGGCTGTGGGCTGTGGGGCGGGG
Genomic   TCACCACTGCCCTGTGGAAGAAGTGGGTGGAGAATACCTCCTCATGCCCA
5'-RACE   TCACCACTGCCCTGTGGAAGAAGTGGGTGGAGAATACCTCCTCATGCCCA
Genomic   AGGCCTTACCTTGGCCTCCCCTCCAGGGCCGGGTGGGGTGGGGCTGGCCA 2β
5'-RACE   AGGCCTTACCTTGGCCTCCCCTCCAGGGCCGGGTGGGGTGGGGCTGGCCA
Genomic   GGGTGTGACCACCGTGCTGGGCAGCAGGCTCCAGTCCCTAACCCCCAGCC
5'-RACE   GGGTGTGACCACCGTGCTGGGCAGCAGGCTCCAGTCCCTAACCCCCAGCC
                          +1
Genomic   ACTACTGGCATGAGGGGTCCCTCAGGGCCCCCAGGCCTCCTCTACGTCCC
5'-RACE   ACTACTGGCATGAGGGGTCCCTCAGGGCCCCCAGGCCTCCTCTACGTCCC
                                                        ⇓
Genomic   ACACCTCCTCCTCTGCCTGCTCTGCCTCCTCCCACCGCCGCTCCAATACG
5'-RACE   ACACCTCCTCCTCTGCCTGCTCTGCCTCCTCCCACCGCCGCTCCAATACG
Genomic   CAGCCGGGCAGAGCCAGATGCCCAAAGACCAGCCCCTGTGGGCACTTCTG
5'-RACE   CAGCCGGGCAGAGCCAGATGCCCAAAGACCAGCCCCTGTGGGCACTTCTG
```

Figure 8.9 Nucleotide sequence of human CRH-R2 5'RACE-derived cDNA and comparison to the corresponding genomic DNA sequence. The 5'UTR of the 5'RACE begins with nt denoted as -226, -201, -97, and -12, upstream from the A of the ATG translation initiation codon in the CRH-R2β exon I, which is denoted as nt +1. The bold letters represent the transcription start site. The underlined sequence corresponds to the previously determined sequence of the CRH-R2β, 2γ and 2α cDNA (Valdenaire *et al.*, 1997, Kostich *et al.*, 1998, Liaw *et al.*, 1996). (⇓) indicates intron positions.

8.4 Discussion

5'RACE with a CRH-R2α exon I primer was used to amplify the 5'-end cDNA for sequencing. The nucleic acid sequence of one fragment revealed that 309 bp fragment corresponded to the CRH-R2β (196 bp) and CRH-R2α (113 bp) exon-I. The PCR product was obtained with a splicing pattern of 2β to 2α (skipping 2γ). The sequence of another fragment contains CRH-R2β (196 bp), CRH-R2γ (227 bp) and CRH-R2α

(113 bp) exon-I. The fragment contains another splicing pattern: 2β to 2γ to 2α. DNA sequencing of both fragments revealed that the 5'end of the CRH-R2 transcript is located at 12 bp upstream of the CRH-R2β of the ATG translation initiation site. On the other hand, 5'RACE analysis extended the previously known exon I sequence of the human CRH-R2α, and established that the 5'-untranslated region of CRH-R2α mRNA in human hippocampus is 52 nt. The 5'RACE cDNA of CRH-R2α sequence corresponded exactly to the genomic DNA sequence.

5'RACE was performed using the CRH-R2γ cDNA specific primer with human hippocampus cDNA. For each cDNA fragment, the nucleotide sequence of six independent clones was determined, which was the same for all clones. DNA sequencing of both fragments revealed that the 5'end of the CRH-R2 transcript is located at 97 bp upstream of the CRH-R2β of the ATG translation initiation site. The nucleic acid sequence of one fragment revealed that 488 bp fragment corresponded to the CRH-R2β (280 bp) and CRH-R2γ (208 bp) specific exon-I. The sequence of another fragment contains CRH-R2β (280 bp), CRH-R2γ (208 bp) specific exons and an extra 195 bp intronic sequence. Alternative splicing in the N-terminus of the CRH-R2 can generate an extra 195 bp. The cDNA inserted at between the CRH-R2β Ib and CRH-R2γ exons, potentially encodes a truncated protein as a result of an in-frame termination condon in intron I. The potential physiological significance of such a truncated receptor is not currently clear. The corresponding mRNA cannot be translated into functional receptor, since this sequence contains in-frame termination codon. Moore *et al.,* 1995 have identified a third isoform of the human calcitonin receptor, which can be generated by alternative splicing in the first intracellular loop. The presence of this alternatively spliced exon would result in the addition of six amino acids at the end of the first transmembrane domain, followed by the premature termination of translation (Moore *et al.,*. 1995). Recently, Miyata *et al.,* have identified a novel isoform of the CRH receptor from the rat amygdala cDNA library. One of them contains intron 6, and the other contains both intron 6 and 7 of CRH-R2α genomic DNA. These two clones encode an identical truncated protein, which is composed of 236 amino acids, because an in-frame stop codon is located in intron 6 (Miyata *et al.,* 1999). The above described method has also been used for myometrial tissues, but the extra 195 bp intronic sequence has not been detected in human myometrium.

The sequence of the complete 5'-UTR of the human CRH-R2β mRNA is not known. The nucleotide sequence previously determined for the human CRH-R2β cDNA included a sequence of 30 nt flanking the ATG translation start site (Valdenaire *et al.*, 1997). We have extended this 5'-UTR using 5'RACE. The long 5'RACE derived cDNA clones included the previously reported 30 nt 5'-UTR sequence upstream of the CRH-R2β open reading frame as well as an additional 196 nt of 5'-UTR sequence. The short 5'RACE derived cDNA clones included the previously reported 30 nt 5'-UTR sequence upstream of the CRH-R2β open reading frame as well as an additional 171 nt of 5'-UTR sequence. Therefore, the 5'-UTR of the human CRH-R2β cDNA are 226 and 201 nt.

Individual genes can give rise to multiple size of transcripts and hence to multiple 5'RACE fragments via at least two mechanisms: (1) Alternative splicing can cause multiple products in 5'RACE. (2) Use of different transcription start sites causes multiple 5'RACE products. The CRH-R2 gene is most likely due to both reasons. We have determined the structure of the 5'end of the CRH-R2 gene that is organised with the 5' alternative exons. When 5'RACE with CRH-R2β and CRH-R2α primers were used to identify the transcriptional start sites of the CRH-R2 gene, each two different cDNA fragments were generated.

In summary, the 5'RACE analysis identified four transcriptional start sites located at 226, 201, 97, and 12 bp upstream of the translation initiation site of the CRH-R2β. In TATA-less genes, the mechanism of transcriptional initiation and its regulation are not uniformly established. In general, GC-rich domains and initiator elements have been proposed to act co-operatively to direct gene transcription (Smale and Baltimore, 1989). GC-rich promoters, found primarily in housekeeping genes, usually contain several transcription start sites spread over a fairly large region and several potential binding sites for the transcription factor Sp1. The CRH-R2 prompter most likely belongs to this class. Other TATA-less promoters are not GC-rich and initiate transcription at only one or a few tightly clustered start sites (Sehgal *et al.*, 1988). Many of these latter types of promoters, including the promoter for the terminal

deoxynucleotidyl transferase gene, are regulated during differentiation or development.

On the other hand, it is probable that there is a single start site at 226 bp. It is more likely that the 12, 97, and 201 bp are incomplete transcripts. There are several possible sources of incomplete fragments, which are generated from correctly primed sites. Premature termination of first strand cDNA synthesis caused by pausing of reverse transcription generally causes multiple 5'RACE products. This is a common problem with larger RNAs, and it is a difficult problem to overcome, since it is due to an intrinsic limitation of reverse transcription. Difficulty on amplifying certain "difficult" genes can cause multiple products in 5'RACE, often a result of high GC contents. The promoter of the human CRH-R2 gene is a GC-rich (61%) domain. Further studies will be required to confirm the transcriptional start sites of the CRH-R2 gene. The 5'RACE method can be employed using other tissues, as well as using S1nuclease mapping, primer extension, and RNase protection assays methods.

The CRH-R2α, 2γ of 5'RACE derived cDNA clones included the 5'-Untranslated region (UTR) upstream of the CRH-R2β open reading frame. The results suggest that there may be a single promoter regulating the expression of all (2α, 2β, and 2γ) subtypes. The promoter of CRH-R2 lies upstream of 5'-end of the CRH-R2β.

Alternative splicing is a widespread mechanism for producing variant forms of proteins from a single gene. We have determined the structure of the 5'end of the CRH-R2 gene which is organised with the 5' alternative exons arranged as CRH-R2β-2γ-2α. Structural comparison of these CRH-R2 subtypes showed that they differ only in their N terminus; the 34 amino acids N terminal to CRH-R2α are replaced by a 61amino acid sequence to form the CRH-R2β or a 20 amino acid sequence to form the CRH-R2γ. The mature CRH-R2 subtypes mRNA appears to be generated via 5' exon differential splicing. The results conform that CRH-R2 exists in the three splice variant forms.

Comparison of the genomic sequence with cDNA obtained by RT-PCR and 5'RACE established the organisation of the 5'end of the CRH-R2 gene. The 5'end of the CRH-

R2 (2β, 2γ, and 2α exon-1) gene is composed of 4 exons and 3 introns distributed over 18 kb. The 2β exon-1 is interrupted by a large intron (10659 bp) in the 5'end of CRH-R2 gene. The CRH-R2β and CRH-R2γ exons possess good consensus splice junctions flanking an apparent 2575 bp intron. The distance between the CRH-R2γ-exon and the CRH-R2α-exon is 4068 bp (see Chapter 7). Accurate estimates of the distance between each exon and the composition of each exon was based on a combination of Southern analysis, DNA sequencing and PCR amplification. The 5'end of the CRH-R2 (2β, 2γ, and 2α exon-I) cDNA sequence corresponded exactly to the genomic DNA sequence.

The CRH-R gene exhibits substantial similarity to the calcitonin/PTH receptor family that is characterised by the presence of introns within its transmembrane/cytoplasmic module, highly conserved cysteines in its extracellular domain, and a highly conserved first intracellular loop (Kong *et al.*, 1994). Intron-exon junctions conform to the consensus sequences, with exons flanked by AG and GT at 5' and 3' ends, respectively. Elucidation of the gene structure of the porcine calcitonin receptor in 1994 (Zolnierowicz *et al.*, 1994) first revealed the unique intron-exon organisation of a receptor in the group of G protein-coupled receptors. The intron-exon organisation of the 5'end of the CRH-R1 (see Chapter 5) and CRH-R2 gene is strikingly similar to that of genes encoding other members of the group II family of G protein-coupled receptors (*i.e.* the secretin/calcitonin/PTH receptor family).

Chapter 9 Construction of expression vector for transient transfection of eukaryotic cells

9.1 Introduction

A variety of reporter gene assay systems are available to investigate transcriptional gene control. Regions of the distal and proximal promoter that are important for basal or regulated expression can be deduced by deleting or mutating putative control regions and observing the effect of such deletions on reporter gene activity. To determine whether a DNA sequence is both necessary and sufficient for either cell-specific or regulated expression, it can be fused in single or multiple copies upstream of a heterologous promoter; and the activity of this promoter can be measured using a reporter gene assay system. The reporter gene constructs can also be used to investigate protein/DNA interactions which regulate gene expression.

A standard first approach is to perform a series of deletion mutants, generally making progressive truncations from the 5' end of the promoter. Such mutations can be made easily using convenient restriction sites, by digesting with a progressive nuclease such as ExoIII nuclease, or by using PCR to amplify different fragments of the promoter. Although this technique can reveal promoter regions which are important for activation or repression, it should be recognised that removal of elements may preclude subsequent studies of interactions between different regions of the promoter. A second and often subsequent strategy is to perform block mutations or point mutations in specific regions of the promoter. Frequently, these regions are identified based upon the earlier deletion studies or based upon candidate sites identified by computer or by inspection. Block type mutants can be inserted using "linker-scanning mutagenesis" or by PCR. Point mutations are generally introduced using various site-directed mutagenesis protocols.

Reporter gene systems are designed for functional testing of eukaryotic transcriptional regulatory sequences. These systems are mammalian plasmid expression vectors containing several components. They include prokaryotic sequences that permit propagation and growth in *E.coli,* antibiotic resistance to facilitate recombinant

plasmid selection and a number of unique restriction endonuclease sites for insertion of the DNA sequences of interest. The reporter gene transcriptional unit comprises the reporter gene sequences, which are responsible for driving its expression, intron sequences and mRNA polyadenylation signals. The most commonly used reporter genes are those encoding enzymes such as firefly luciferase and chloramphenicol acetyl transferase (CAT). The latter has low background activity in mammalian cells. Eukaryotic cells never express this enzyme, but the CAT gene can be expressed from eukaryotic promoters (Gorman *et al.*, 1982).

The β-galactosidase enzyme is another bacterial product that can be easily measured. Although less sensitive than luciferase or CAT, it has the additional advantage that it can be detected easily in histologic samples. Thus, β-galactosidase is used frequently in transgenic studies. Recently, the green fluorescent protein, derived from jellyfish, has provided anther useful tool for examining expression in transfected cells or in various tissues. It has the advantage of spontaneous fluorescence, even allowing studies in living cells and animals.

The characterisation of transcriptional regulatory sequences at the 5'flanking region of the CRH-R2 gene included functional studies using CAT reporter gene systems. The general scheme for such studies involved modification of native CRH-R2 promoter sequences by deletion mutations. Reporter gene constructs were transiently transfected into HEK 293 cells for functional studies.

9.2 pCAT3-Basic constructs

9.2.1 Preparation of the CRH-R2 promoter cloning fragments

The pCAT 3-Basic plasmid (Promega) was used as the cloning vector. This vector lacks eukaryotic promoter and enhancer sequences. Expression of CAT activity in cells transfected with this plasmid depends upon the insertion of a functional promoter upstream of the CAT gene. Circular map showing the multiple cloning sites is provided in Figure 9.1.

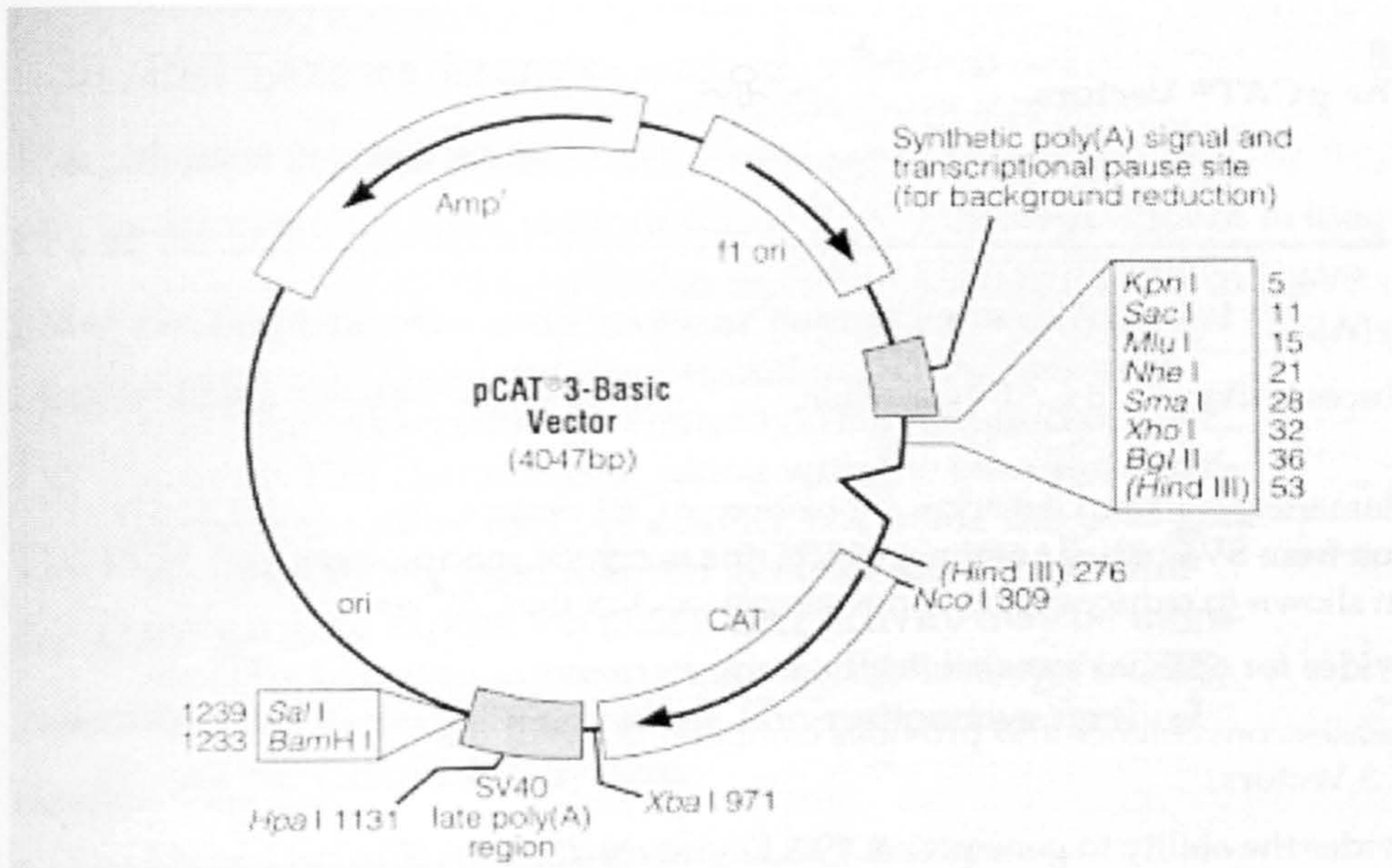


Figure 9.1 pCAT3-Basic vector circle map and sequence reference points.

Sequence reference points:

a. SV40 late poly (A) region	1001-1222
b. CAT gene	311-967
c. upstream poly (A) region	3887-4040
d. Multiple cloning site (*Kpn*I – *Bgl* II)	1-58
e. RV primer 3 binding site	3989-4008
f. RV primer 4 binding site	1309-1290
g. β-lactamase gene (Amp^r)	3169-2312
h. f1 origin	3302-3756

A CAT reporter system (Promega) was used with the enzyme-linked immunosorbert assay (ELISA) method for monitoring CAT enzyme activity in transiently transfected cells (see Chapter 10).

The polymerase chain reaction was employed to generate a 1457 bp fragment (named PF) of the DNA corresponding to the 5'proximal region, the exon I of the CRH-R2 gene. The fragment was subcloned and sequenced. Sequencing analysis confirmed the identity of the amplified DNA.

The PCR fragments spanning different regions of promoter were cloned upstream of the CAT reporter gene in the promega pCAT3-Basic vector. These inserts were generated from genomic DNA using PF as a template and PCR reaction with a high fidelity Advantage 2 Polymerase (Clontech). The oligonucleotides used for this PCR have a restriction site 5' to facilitate cloning into the pCAT3- Basic vector polylinker. For constructs derived from the putative promoter region of CRH-R2, a common 3' reverse oligonucleotide was used in combination with different 5' oligonucletides (see Table 9.1).

Table 9.1
Primers for generating deletion constructs

Nucleotide number (5' to 3')	Sequence	Construt name
-1393 to -1371	CCGGTACCGACCTGTGAGATTTCACAGGAG	P1
-1035 to 1014	CCGGTACCGAGGCTCTGGAAGAACAGACC	P2
-553 to --532	CCGGTACCACCTACCCACAGTACAGTATC KpnI	P3
(5' to 3') 47 to 26	CGGCTCGAGAGGTGTGGGACGTAGAGGAGG XhoI	

The PCR amplifications were carried out in a total volume of 50µl of 40 mM Tricine-KoH pH 9.2, 15 mm KOAc, 3.5 Mg $(OAC)_2$, 0.2 mM each dNTPs, 0.5 µM of each upstream and downstream primer and Advantage 2 Polymerase (Clontech). After an initial denaturation at 95° C for 2 minutes, 5 cycles of PCR were carried out, followed by denaturation at 95°C for 30 second and annealing/extension at 72°C for 1.5 minutes. This was followed by 5 cycles of PCR involving denaturation at 95°C for 30 second and annealing/extension at 70°C for 1.5minutes. A further 30 cycles of PCR was carried out with denaturation at 95°C for 30 second, annealing/extension at 68°C for 1.5 minutes. The final extension was at 72°C for 3 minutes.

The PCR products were electrophoresed on a 1.0% agarose gel as shown in Figure 9.2 and purified from the gel with a gel extraction kit (Qiagen). The PCR products were sequenced to confirm their identity by the dideoxy chain termination method using an Applied Biosystems 373A automated DNA sequencer. The PCR fragments were digested with *Kpn* I and *Xho* I and ligated into the equivalent site of the pCAT3-Basic vector.

9.2.2 Preparation of the cloning vector pCAT3-Basic

The original vector pCAT3-Basic was subjected to a double digest with the restriction endonucleases *Kpn* I and *Xho* I and electrophoresed on a 1.0% agarose gel and purified from the gel with a gel extraction kit (Qiagen). The fragment pCAT3-basic (*Kpn I/Xho* I) was used to prepare the expression vectors P1, P2, and P3.

9.3 Ligated into equivalent site of the pCAT3-Basic

Each double digest of PCR fragment was incubated with 0.1µg pCAT3-Basic plasmid *(Kpn* I and *Xho* I) at approximately equimolar quantity and ligated under the catalysis of T4 DNA ligase. Ligation mixtures were transformed into competent DH5α cells and recombinants were selected overnight on agar plates containing ampicillin.

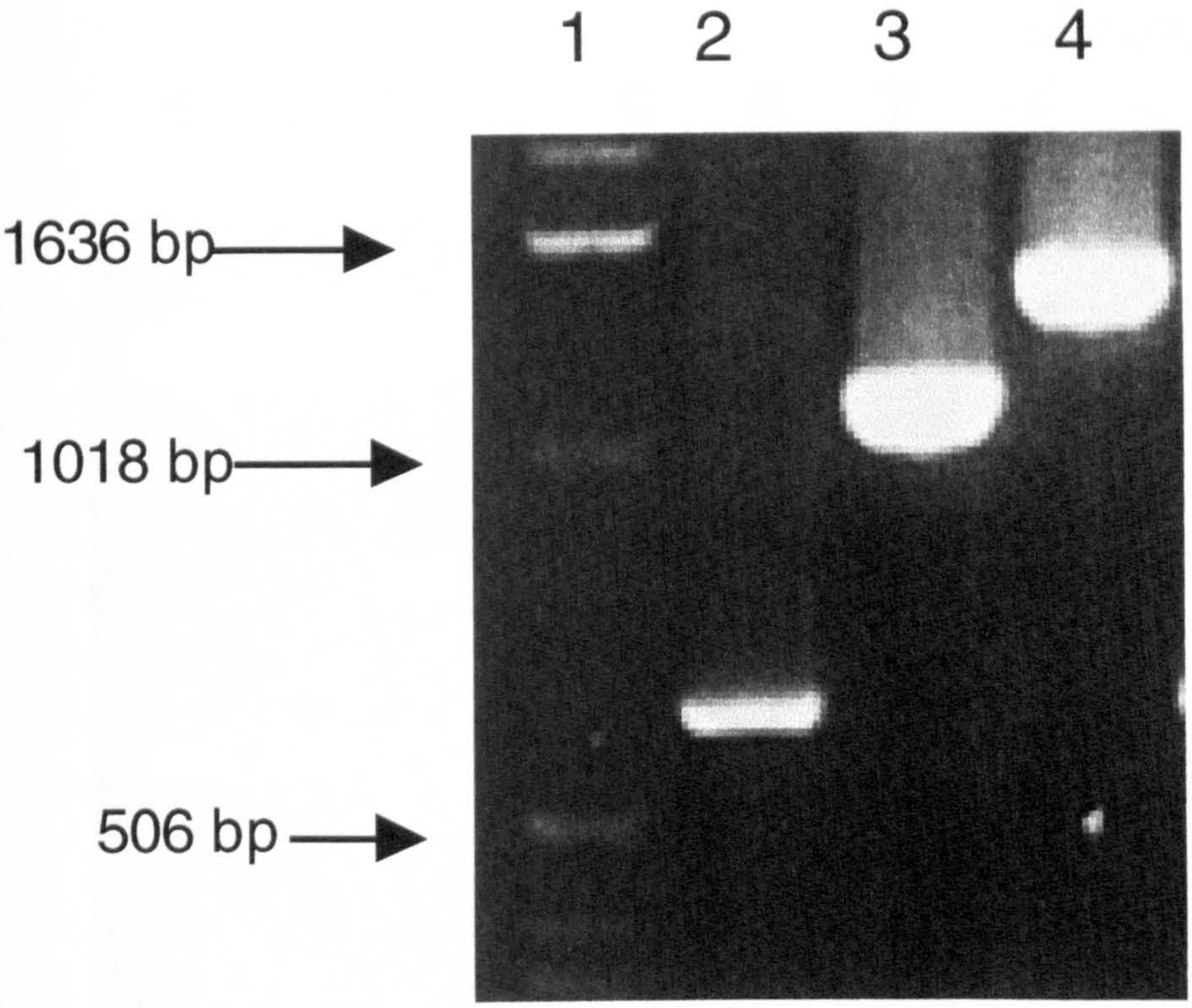


Figure 9.2 Amplified different fragments of the promoter region using PCR
Lane 1 is a DNA marker. Lane 2 shows PCR amplification of a 600 bp fragment (P3), which contained 553 bp upstream of the translation initiation site of the CRH-R2β and an 47 bp downstream sequence. Lane 3 shows PCR amplification of a 1082 bp fragment (P2), which contained 1035 bp upstream of the translation initiation site of the CRH-R2β and an 47 bp downstream sequence. Lane 4 shows PCR amplification of a 1440 bp fragment (P1), which contained 1393 bp upstream of the translation initiation site of the CRH-R2β and an 47 bp downstream sequence.

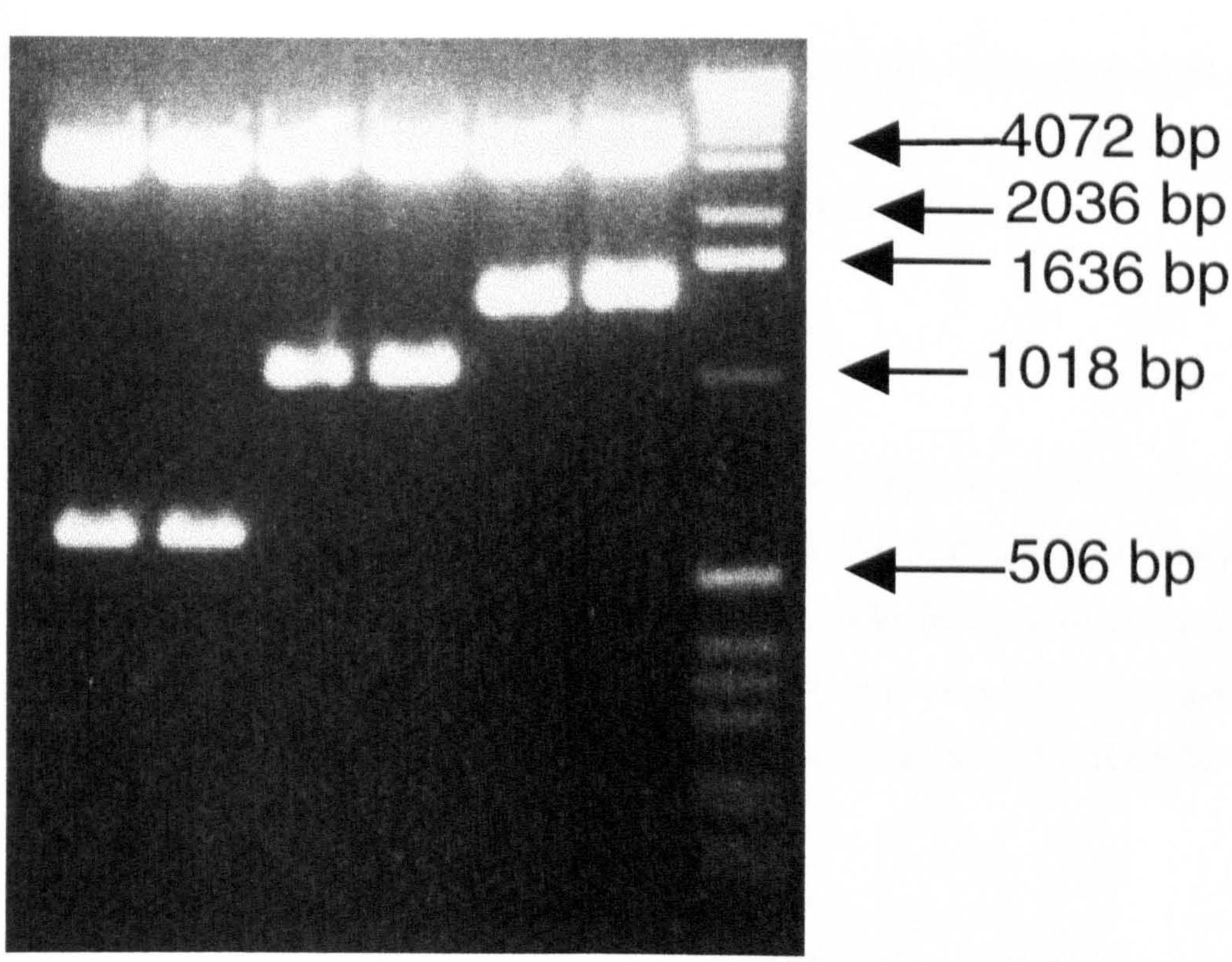


Figure 9.3 Analysis result of plsmid DNA for the insert by restriction enzyme digestion using *Kpn* I and *Xho* I.

Two bands were visualised on an ethidium bromide stained agarose gel that corresponds to the pCAT3-Basic plasmid (4047 bp) and the insert. Lane 1 and 2 show P3 fragment. Lane 3 and 4 show P2 fragment. Lane 5 and 6 show P1 fragment. Lane 7 are DNA markers.

Plasmid DNA was isolated from overnight culture derived from the single colonies on the agar plates. Plasmid DNA was digested using *Kpn* I and *Xho* I enzymes. Two bands were visualised on an ethidium bromide stained agarose gel corresponding to the plasmid and the insert (Figure 9.3). The inserts were sequenced using an automated DNA sequencer.

All DNA plasmid used in transfections was purified using QIAGEN plasmid Maxi kit and analysed by agarose gel electrophoresis to verify plasmid integrity. Plasmid DNA was determined using an U V spectrophotometer (260 and 280 nM).

9.4 Discussion

For the purposes of the functional study of the 5'-flanking region of the CRH-R2 gene, CAT expression vector containing parts of the CRH-R2 proximal promoter was constructed. A set of deletion constructs was generated. The deletion constructs were cloned just from 5' to the CAT cDNA and the length (upstream of the translation initiation site) of the CRH-R2 5' sequence inserted are 1393 bp (P1), 1035 bp (P2), and 553 bp (P3).

One functional type of cloning vector, pCAT3-Basic, was employed to perform cloning and to demonstrate the presence of CRH-R2 regulatory sequences which are capable of initiating constitutive transcription. At the same time, we also studied the transcriptional effects due to additional regulatory sequences present in the inserted DNA fragment. The above described CAT expression vector have been constructed and can be used to study how the putative regulatory elements are responsible for the transcriptional induction of the CRH-R2 gene which are particularly stimulated using the treatments like adding CRH, forskolin, phorbol-myristate-acetate, glucocorticoids or estrogen.

The functional investigation of CRH-R2 gene regulatory sequences has been carried out and reported in this chapter. The CAT reporter gene vector is also constructed for the purpose of determining the transcriptional activity of this promoter, which will be reported in Chapter 10.

Chapter 10 CAT reporter gene analysis of regulation of the CRH-R2 gene

10.1 Introduction

The study of eukaryotic gene regulation and expression has been advanced by technology which allows the introduction of nucleic acids into eukaryotic cells. Physical and chemical methods are used to introduce nucleic acids in a process referred to as transfection. The most widely used procedure is the calcium phosphate-DNA precipitation technique. This procedure is inexpensive, efficiently transfects most transformed monolayer cell lines, and is relatively reproducible. The major technical problem is the exquisite sensitivity to pH. Another widely used protocol involves the use of DEAE-dextran which is particularly useful for cells that tend to grow in suspension. Recently, lipid-mediated or liposome-mediated transfection protocols have gained popularity. These techniques are highly efficient for most cultured mammalian cells. Although commercial sources for these reagents can be expensive, it is possible to prepare effective lipid mixtures from relatively inexpensive reagents (Rose *et al.*, 1991). Electroporation can be used to introduce DNA into most cell types, including primary cell cultures. This procedure uses an electric current to transiently open cell membranes and allows transfering of DNA into the cell. It requires an electroporation device and the cuvettes are relatively expensive. The electroporation conditions (voltage, capacitance) and other parameters for buffer and DNA concentration need to be optimised for each cell type. A number of viral vectors have also been developed for introducing genes into cells. The replication-deficient adenovirus can be used efficiently in a wide variety of cell types. Retroviruses are particularly useful for integrating DNA into cell lines in a stable manner. Finally, DNA can also be microinjected into cells. Although it is useful for certain special applications, in practical terms, only a limited number of cells can be microinjected at any one time. We therefore chose to use a liposome-mediated transfection method.

Several key issues should be considered in the selection of a cell line for transfection experiments. For many genes, expression is limited to a restricted number of tissues

and cell types. Thus, it is critical to choose cell types that normally express the gene if one is to correctly identify critical regulatory elements and their cognate transcription factors. For this reason, cell lines should be screened to ascertain whether they normally express the gene of interest. Another issue concerns how easily the cells can be transfected and how stable the cells are in culture.

To test the promoter activity of the 5'-flanking region of the CRH-R2 gene, two cell types (Myometrial primary cell cultures and HEK-293 cells) were employed.

10.2 Transfection of HEK-293 and Myometrial cell with Lipofectamine reagent

HEK 293 and Myometrial cells were maintained in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum (Gibco-BRL), penicillin (200Units/ml), streptomycin (200mg/ml) and incubated in 5% CO_2 at 37°C. Confluent cell flasks were trypsinized, the number of cells were counted and in six-well plates (Nunclon) were seeded with 3 x 10^5 cell per well. Cells were grown in 2 ml of Dulbecos modified Eagle medium, 10% fatal bovine serum, penicillin (200Units/ml), streptomycin (200mg/ml) and incubated in 5% CO_2 at 37°C for 24-48 hours before the tranfection. All DNA plasmid used in transfection were purified using QIAGEN plasmid Maxi Kit and were analysed by agarose gel electrophoresis to verify plasmid integrity. The pCAT3-Basic-less plasmid was routinely in all transfection experiments as negative controls.

Liposome-mediated transfection was performed using Lipofectamine Reagent (Life Technologies) according to the method of Felgner *et al.* For each deletion construct, 1.0-1.5 µg test DNA was diluted into 100µl of OPTI-MEM serum-free medium (solution A). PLUS Reagent was added to the DNA in solution A and incubated at room temperature for 15 minutes. This solution A was then mixed with lipofectamine (6µl of lipofectamine in 100µl OPTI-MEM serum-free medium solution B). The solutions were combined and incubated for 30 minutes at room temperature. Then 800µl of serum-free medium were added on, and the solution was then carefully dripped onto the cells, which had been washed once with serum-free medium. The

cells were incubated for 3 hours before the transfection solution was removed and replaced by serum-containing medium. After an incubation period of 48 hours, the cells were harvested.

10.3 Determination of CAT activity

10.3.1 Preparation of cell extracts

The wells containing monolayers of transiently transfected cells were washed three times with 3ml of ice cold PBS and the cells were removed with a fine-tipped Pasteur pipette. 340µl of lysis buffer (Boehringer mannheim) was then added and the cells were allowed to stand for 30 minutes at room temperature. The cells were transferred into microcentrifuge tubes and centrifuged at 13,000 rpm for 10 minutes at 4°C. The supernatants were transferred into microcentrifuge tubes. Cell lysis was ensured by three cycles of freeze thawing (2 minutes in ethanol/dry ice then 1 minute in 37°C water bath). The cell debris was subsequently removed by centrifugation at 13,000 rpm for 10 minutes. The cell extracts were transferred into clean tubes and stored at -70°C until use.

10.3.2 CAT enzyme-linked immunosorbent assay

The CAT activity of cytoplasmic extracts from cells transiently transfected with CAT expression vectors was measured by the enzyme-linked immunosorbent assay (ELISA) method.

The CAT activity was measured using a CAT ELISA kit (Boehringer, Mannheim) according to the manufacturer's instruction. The only modification was that 10µl of the cell extract was added into the well instead of the recommended (200µl/well).

Background CAT expression from the vector was determined by transfection of the empty pCAT3-Basic plasmid.

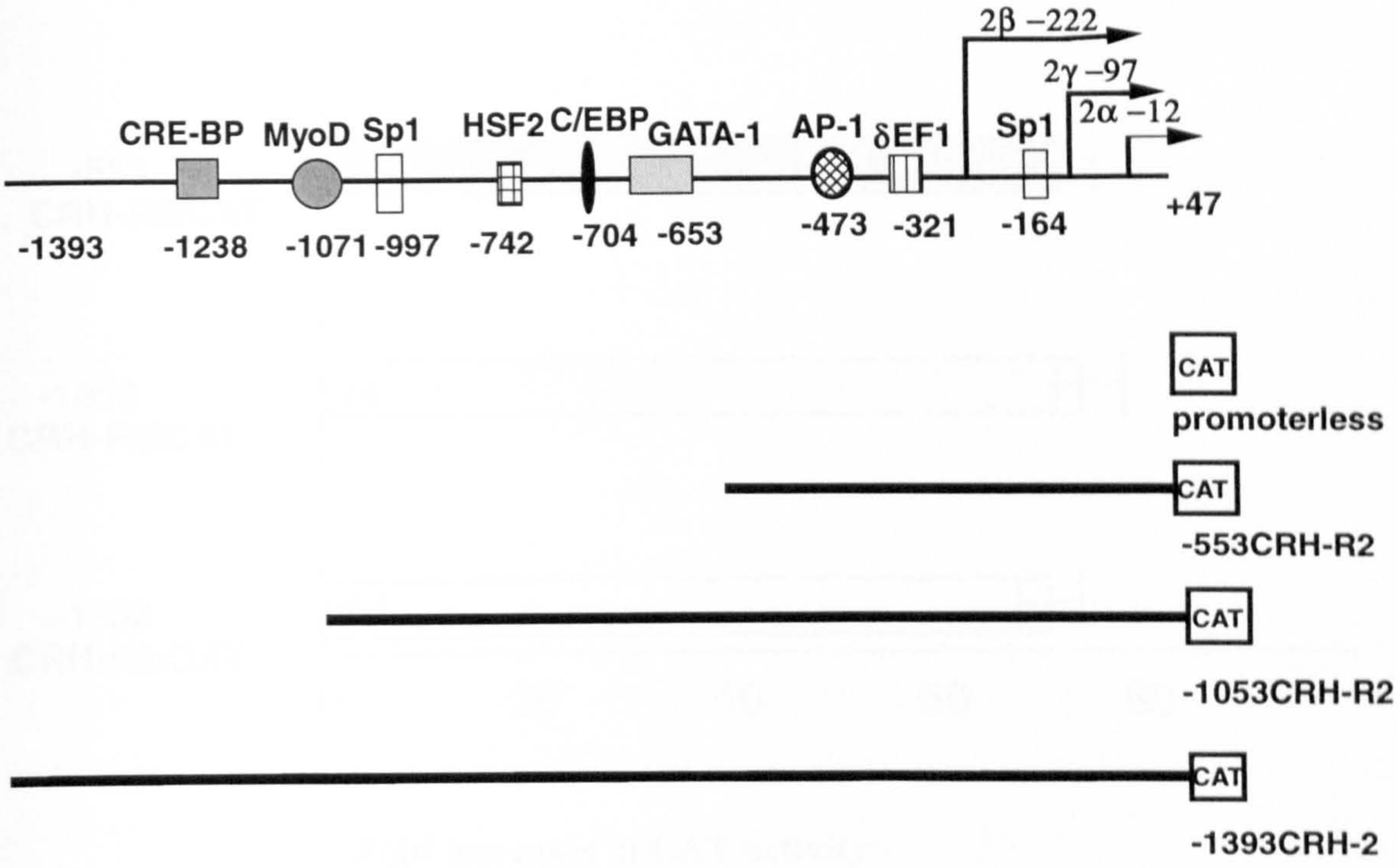
10.4 Transient expression analysis of the 5'-flanking region

To further analyse the 5'-flanking region of the CRH-R2 gene for constitutive promoter activity in HEK 293 cells, various lengths of 5'-flanking regions created by PCR were placed onto upstream of the CAT reporter gene (Figure.10.1). Only background activity was obtained with the promoterless pCAT3-basic CAT vector.

To determine whether the CRH-R2 5'-flanking region can direct cell expression, the proximal 553 bp, 1035 bp and 1393 bp upstream of the translation initiation site of the CRH-R2β were inserted into a transient expression vector, pCAT3-Basic. The pCAT-Basic vector contains CAT as the reporter gene, and the resulting plasmids (-553 CRH-R2/CAT, -1035 CRH-R2/CAT and -1393 CRH-R2/CAT) were transiently transfected into HEK 293 and myometrial primary cell culture. Gene transfer studies were carried out by liposome-mediated transfection and CAT activity was measured by the enzyme-linked immunosorbent assay (ELISA) method.

Our studies showed that 1393bp (P1), 1035bp (P2), and 553bp (P3) of the CRH-R2 promoter directed high levels of CAT expression in HEK 293 cells as compared with the promoterless pCAT3-basic CAT vector. In contrast, we observed no significant activity of any of these constructs in the myometrial cell cultures.

A 72-fold increase in activity was observed with the construct containing 553 bp (p3) of 5'-flanking region (Figure. 10.2). The construct containing 1035 bp (p2) produced a 74-fold increase in activity compared with the promoterless control. Another 67-fold increase in activity was observed with the construct containing 1393 bp (P1) of 5'-flanking region. These results suggest that the regions between + 47 and -553, -553 and -1035 contain important positive regulators of CRH-R expression in HEK 293 cells. The negative control, pCAT3-Basic without any 5'-flanking region, caused very low CAT activity.



hCRH-R2 promoter constructs

Figure 10.1 The deletion constructs of the CRH-R2 5'-flanking region
The schematic diagrams represent a series of CRH-R2 promoter-CAT gene plasmids with variable 5'-end (from -1393 to-553) and the same 3'-end (+47). Each construct was transiently transfected into HEK 293 cells. The transcription factor binding elements are labelled above. Their relative nucleotide positions to ATG are indicated blow. ATG translation initiation codon in exon I, is denoted by nt +1. The arrow indicates the transcription initiation sites.

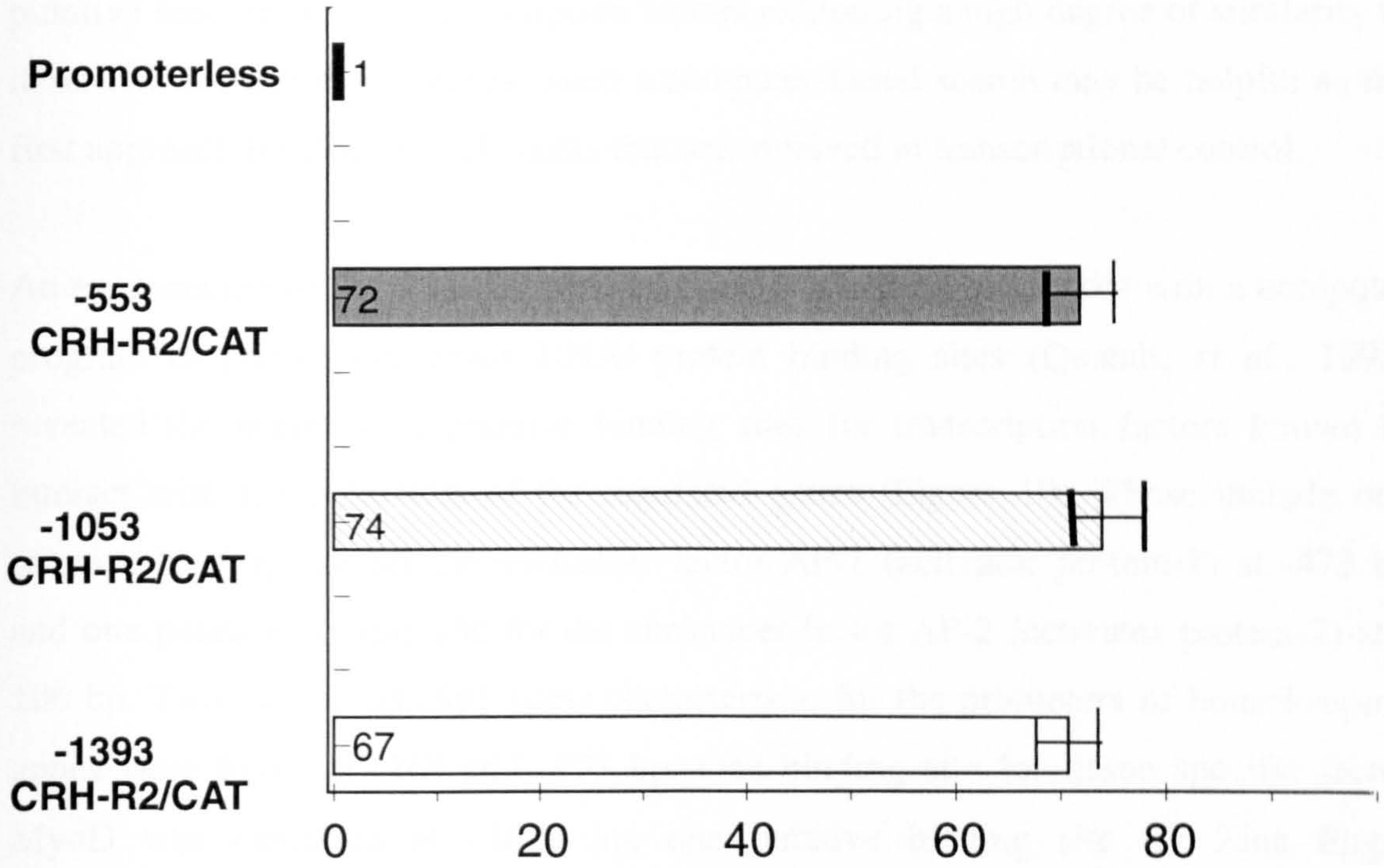


Figure 10.2 Deletion analysis of human CRH-R2 promoter activity.

The average fold increase for each construct is indicated to the right of each bar. CAT activities of the test samples were referenced to those from promoterless pCAT3-Basic. Data are the means ± SEM of at least three independent experiments performed in triplicate.

10.5 Discussion

The inspection of the 1.3 kb 5'-flanking promoter region revealed a broad selection of putative binding sites for transcription factors exhibiting a high degree of similarity to described consensus sequences. Such a computer-based search may be helpful as the first approach for detecting elements that are involved in transcriptional control.

An examination of the CRH-R2 promoter and 5'-flanking sequences with a computer program to predict consensus DNA/ protein binding sites (Quandt, *et al.*, 1995) revealed the presence of putative binding sites for transcription factors known to interact with the promoters of the regulated genes (Figure 10). These include one putative binding site for the enchancer factor AP-1 (activator protein-1) at -473 bp and one putative binding site for the enchancer factor AP-2 (activator protein-2) at -106 bp. Two GC boxes (Sp1 sites) characteristic for the promoters of housekeeping genes were found at -164 and -997 bp. One binding site for tissue specific factor MyoD was identified at –1079 bp, one putative binding site for Zinc finger transcription for GATA-1 was found at –653 bp and one CCAAT/enhancer binding protein (C/EBP) was found at –704 bp. The CRH-R2 promoter region up to -1393 bp contains sites for GATAs, HSF2, deltaEF1, Ap4, P300, Oct-1 and c-Myb.

In addition, the promoter region contains consensus motifs corresponding to inducible promoter elements that are known to bind transcription factors induced by exogenous stimuli. These include binding sites for the transcription factors CRE-BP (cAMP-response element (CRE)-binding protein) at bp -1238, GRE at –665, SF-1 at –1204 bp.

The pCAT-basic promoter-less plasmid containing the chloramphenical acetyltransferase (CAT) gene was used for construction of reporter gene plasmid to analyse the CRH-R2 gene promoter function. An analysis of promoter and CAT constructs containing 5'-flanking fragments of different length up to -1393 bp by cell transfection revealed the first insight into the functionality and architecture of the promoter of the CRH-R2 gene. In HEK 293 cells, the constructs -553 bp (P3) and -1035 bp (P2) showed an approximately 70-fold increase in promoter activity compared to promoterless pCAT3-Basic. These results show that several potential

positive regulatory elements (possibly two Sp1 and one AP1) are located between -164 and -977 nt. This region contains two Sp-1 consensus binding sites, which may be critical for reporter gene expression. Sp-1 binding sites are known to be promoter-specific activation sequences of RNA polymerase II transcription. The construct -1393 bp (P1) showed approximately 10% diminished activity relative to P2 or P3, which possibly indicated that the sequence from -1035 to -1393 may contain the negative regulatory elements. In contrast, we observed no significant promoter activity of -553 CRH-R2/CAT, -1035 CRH-R2/CAT and -1393 CRH-R2/CAT in myometrial cells. The reason for that may be because of transfection problems.

In addition, transfection of promoter deletion constructs led to the identification of stretches within the promoter necessary for maximal basal activity. Further studies are required to identify the basal level CRH-R2 promoter activity.

Deletion analysis of the promoter region revealed that several putative Sp1, AP-1, and AP-2 binding sites may be important for the basic promoter activity. Furthermore, it is of interest that there were consensus binding sites for MyoD (a transcriptionl regulator in myoblasts and skeletal muscle), GRE, and SF-1 in the strong positive regulatory region. Further studies are required to determine whether these binding sites are functional.

The first 1.3 kb of the 5'-flanking sequence has many characteristics of the housekeeping gene promoter including the presence of Sp1 binding sites and the absence of a consensus TATA box. We demonstrated that this promoter initiates transcription from at least 4 start sites over a 214 bp region. Moreover, the 5'-flanking region contains consensus-binding sites for several transcription factors that may be functionally important, including AP1, CRE-BP, AP2, GRE and SF-1. From the results of the present study, therefore, we speculated that the strong and housekeeping gene-like activity of the CRH-R2 gene promoter might contribute to the ubiquitous expression of the CRH-R2 gene. Furthermore, protein-binding studies such as footprinting or gel shift analysis will help us to verify the involvement of the designated putative sites in the transcriptional control of the CRH-R2 gene.

In summary, our initial characterisation of the promoter region of the human CRH-R2 gene demonstrates that this gene contains a TATA-less promoter region with activity in an HEK-293 cell line. A number of transcription factor-binding sites were identified by sequence homology. Furthermore, characterization of these and other regulatory sites using DNase I protection analysis, gel mobility shift assays, and site-directed mutagenesis will identify which of these putative sites are important for basal CRH-R2 gene expression and which sites play a role in CRH-R2 induction by various extracellular signals.

Chapter 11 Summary of the results and suggestions for future work

11.1 Amplifying and cloning CRH-R subtypes from human myometrium

Our results demonstrated the presence of multiple CRH-R mRNAs in the human myometrium. Six subtypes of the CRH receptor, 1α, 1β, 1c, 2α, 2β, and 2γ, were found in the pregnant human myometrium at term before the onset of labor, whereas only three subtypes, 1α, 1β, and 2β were found in the non-pregnant myometrium. This finding demonstrates, for the first time, that during pregnancy, there is an alteration in the pattern of CRH-R subtype expression. This original observation coupled with our laboratory previous studies (Hillhouse *et al.*, 1993. Grammatopoulos *et al.*, 1995, 1996) argue for a functional role for CRH and/or related peptides in myometrial function. Of particular interest is the finding that the CRH-R2α and CRH-R2γ subtypes are present in human pregnant myometrium. This is the first report of the expression of the CRH-R2γ gene in a human peripheral tissue. The location of this subtype was previously thought to be limited to the brain only. Multiple CRH-R mRNAs have been identified in human myometrium with different expression pattern during pregnancy, which argues for multiple roles for CRH and/or related peptides in myometrial function and suggests distinct functional roles for each receptor during pregnancy. Although the physiological role of CRH in myometrial function is not fully understood, the pattern of expression of the CRH-R mRNA suggests that this receptor have a physiological role in the regulation of uterine function. Further studies are currently under way in order to identify the distribution of CRH receptor subtypes in uterine feto-maternal regions as well as to investigate the signals that regulate the pattern of expression of the CRH-R subtypes.

A novel CRH-R2β spliced variant has been isolated from human myometrium by employing the RT-PCR technique in this project. This variant of the CRH-R2β contains a 315 bp deletion in the exon 3-5. This mutant lacks the entire first and second putative transmembrane domain. This variant of the CRH-R2β took place just at the splicing site, suggesting that it may be due to aberrant splicing event. To date, nothing is known about the

function properties of the mutant receptor protein. It is possible that the CRH-R2β variant may modulate the function of the CRH-R2β. Further studies will be required to determine whether the CRH-R2β variant activates the signalling pathway coupled to adenylate cyclase. Alternative splicing is a widespread mechanism for producing variant forms of proteins from a single gene. Alternative splicing is responsible for the generation of a number of G protein-coupled receptor isoforms. Since alternative splicing is generally believed that the transmembrane domains and the intracellular loops contribute to signal transduction (Probst *et al.*, 1992), alternative splicing events in these regions might give rise to receptor isoforms with different signaling properties. The existence of alternative splicing events indicates that there may be other unknown, splice variants of CRH-R.

11.2 Clone and characterize the promoter region of the human CRH-R2 gene

The genomic organisation of human CRH-R2 alternative exons has been determined in the project. The genomic orders of the alternative exons to be CRH-R2β-2γ-2α. The 5'end of the CRH-R2 (2β, 2γ, and 2α exon I) gene is composed of four exons and three introns distributed over 18 kb. Accurate estimates of the distance between each exon and the composition of each exon was based on a combination of DNA sequencing, PCR amplification and Southern analysis. The CRH-R2β exon I is interrupted by a large intron (10659 bp) in the 5'end of CRH-R2 gene. The CRH-R2β Ib and CRH-R2γ I exons possess good consensus splice junctions flanking an apparent 2575 bp intron. The distance between the CRH-R2γ exon I and the CRH-R2α I exon is 4068 bp. Exon-intron junctions conform with the consensus sequences, with exons flanked by AG and GT at 5' and 3' ends, respectively. The CRH-R2β, 2α, and 2γ cDNA sequences corresponded exactly to the genomic DNA sequence. In addition, between CRH-R2β Ia and Ib intron sequence data (10659 bp) was analyzed using ORF (open reading frame) Finder database searches from the National Center for Biotechnology Information (NCBI). There are numerous ATG codons in between CRH-R2β Ia and Ib intron sequence leading to thirty-two short open reading frames (ORFs). ORF frames have the range from 102 bp to 533bp. However, amplification of this region from cDNA obtained from some tissues

(myometrium, hippocampus) has not indicated the presence of other exons in this region. There are numerous ATG codons in the CRH-R2β Ib intron sequence (2575 bp) which led to seven short open reading frames (ORFs). ORF frames have the range from 102 bp to 366 bp. Using gene specific primers for RT-PCR and 5'RACE, one cDNA fragment of 195 bp was amplified from hippocampus but not myometrial biopsies in the CRH-R2β Ib intron. An extra 195 bp can be generated by alternative splicing in the N-terminals of the CRH-R2. The corresponding mRNA cannot be translated into functional receptor, since this sequence contains in-frame termination codon. The CRH-R2β Ib intron was located in this region for the 5'end of CRH-R2 gene, leading to the possibility of other exons with this region. The potential physiological significance of such a truncated receptor is not currently clear. This finding remains to be further investigated, such as isolation of full-length receptor, examination of the expression of this truncated receptor in the different tissues and cAMP accumulation assay.

A number of transcription start sites exist (typically for a TATA-less CG-rich promoter) and the promoter of the human CRH-R2 most likely belongs to this class. By the 5'-rapid amplification of cDNA end, the transcription start sites were mapped to -256, -201, -97, and -12 bases upstream of the ATG translation initiation codon of the CRH-R2β. This is a hypothesis or prediction only and further experiments may be required for providing more solid evidence. On the other hand, it is probable that there is a single start site at 226 bp. It is more likely that the 12, 97, and 201 bp are incomplete transcripts. Further studies will be required to confirm the transcriptional start sites of the CRH-R2 gene. The 5'RACE method can be employed using other tissues, as well as using S1nuclease mapping, primer extension, and RNase protection assays methods.

We have cloned the 5'-flanking region of the CRH-R2 gene by the PCR based genomic walking method. CRH-R2β specific primer was designed, based on the published CRH-R2β cDNA (Valdenaire *et al.*, 1997) and the 5'RACE-derived cDNA sequence. We amplified, cloned and sequenced the longer 1393 base pair fragment, which included the previously reported 30 nt 5'-UTR sequence upstream of the CRH-R2β open. Alignment of the 5'-flanking sequence with the 5'RACE derived cDNA sequence of the CRH-R2β

revealed that the DNA fragment included the 226 nt 5'-UTR sequence upstream of the CRH-R2β open reading frame. The promoter of CRH-R2 lies upstream of the 5'-end of the CRH-R2β. The results suggest that there may be a single promoter regulating the expression of all (2α, 2β, and 2γ) subtypes. Further studies will be required to examine whether the CRH-R2 transcripts are expressed in a tissue-specific fashion, as well as those found in the rodent system.

For functional analysis, 5'-flanking sequence up to -1393 bp were used to the Chloramphenicol acetyltransferase (CAT) gene and tested using a HEK-293 cell line by transfection CAT assays. An analysis of promoter and CAT constructs containing 5'-flanking fragments of different lengths up to -1393 bp by cell transfection revealed the first insight into the functionality and architecture of the promoter of the CRH-R2 gene. In HKT 293 cells, the constructs -553 bp (P3) and -1035 bp (P2) showed an approximately 70-fold increase in promoter activity compared to promoterless pCAT3-Basic. These results showed that several potential positive regulatory elements (possibly Sp1, AP-1 AP-2) are located at between -164 and -977 nt. This region contains two Sp-1 consensus-binding sites, which may be critical for report gene expression. Sp-1 binding sites are known to be promoter-specific activation sequences of RNA polymerase II transcription. The results confirm that analyzed gene is the actively transcribed CRH-R2 gene. Further studies are required to identify the basal level CRH-R2 promoter activity. The hormonal regulation of the CRH-R2 promoter must be identified by further analysis. It is also required to analyze the effect of various agents on -1393 bp of CRH-R2 promoter in HEK 293 cells, such as forskolin, phorbol-myristate-acetate.

The 5'-flanking region of the human CRH-R2 gene was determined. The promoter region lacks a typical TATA or CCAAT box. There are potential *cis*-acting elements (SP1, AP1, AP2, GATA, CRE-BP and MyoD) in the 5'-flanking sequence. The CRH-R2 5'-flanking sequence has many characteristics of a housekeeping gene promoter including its residence in a GC-rich region, contain several transcription start sites spread over a fairly large region, several potential binding sites for the transcription factor Sp1 and the absence of a consensus TATA box. The absence of TATA and CAAT box motifs is a

feature shared with other related family II GPCR genes. Therefore, we speculated that the strong and housekeeping gene-like activity of the CRH-R2 gene promoter might contribute to the ubiquitous expression of the CRH-R2 gene.

The promoter region of eukaryotic genes are generally composed of multiple binding sites for transcriptional activators and repressors that act in combination to regulate expression of linked gene (Roesler and Park 1998). Our initial characterization of the promoter region of the human CRH-R2 gene demonstrates that this gene contains a TATA-less promoter region with activity in an HEK-293 cell line. Comparison of the 5'-flanking sequence of the CRH-R2 gene revealed the existence of several transcription factor-binding sites. Further, characterization of these and other regulatory sites using DNase I protection analysis, gel mobility shift assays, and site-directed mutagenesis will identify which of these putative sites are important for basal CRH-R2 gene expression and which sites play a role in CRH-R2 induction by various extracellular signals. In addition, clarifying the role of CRH-R2 gene regulation should facilitate further studies on delineating the mechanisms of hormone regulation of the promoter region.

Reference:

Abou-Samra, A., Harwood, J. P., Catt, K. J., Aguilera, G. (1987). Mechanisms of action of CRF and other regulators of ACTH release in pituitary corticotrophins. *Ann New York Acad Sci.* **512,** 67-84.

Aguilera, G. (1998). Corticotrophin releasing hormone, receptor regulation and the stress response. *Trends in Endocrinology and Metabolism.* **9,** 329-336.

Albanese, C., Kay, T. W. H., Troccoli, N. M., Jameson, J. L. (1991). Novel cyclic adenosine 3',5'-monophosphate response element in the human chorionic gonadotropin β-subunit gene. *Mol. Endocrinol.* **5,** 693-702.

Albrandt, K., Brady, E. M., Moore, C. X., Mull, E., Sierzega, M. E., Beaumont, K. (1995). Molecular cloning and functional expression of a third isoform of the human calcition receptor and partial characterization of the calcitonin receptor gene. *Endocrinology.* **136,** 5377-5384.

Andrews, S., Langston, Y., Stoneking, T., Maupin, R. (1998). The sequence of Homo sapiens PAC clone RP5-1143H19. *Unpublished* GenBank accession No. AC004976.

Angel, P., Imagawa, M., and Chiu, R. (1987). Phorbol ester-inducible genes contain a common cis element recognized by a TPA-modulated trans-acting factor. *Cell.***49,** 729-739.

Arimura, A., Shioda, S., (1995). Pituitary adenylate cyclase activating polypeptide (PACAP) and its receptors: neuroendocrine and endocrine interaction. *Frontiers Neuroendocrinol.* **16,** 53-88.

Battaglia, G., Webster, E. L., and De Souza, E. B., (1987). Characterization of corticotronpin releasing factor receptor-mediated adenylate cyclase activity in the rat central nervous system. *Synpse (NY).* **1,** 572-581.

Behan, D. P., Potter, E., Sutton, S., Fischer, W., Lowry, P. J., Vale, W. W. (1993). Corticotropin-releasing factor-binding protein: A putative peripheral and central modulator of the CRF family of neuropeptides. *Ann. N Y Acad. Sci.* **697,** 1-8.

Behan, D. P., Khongsaly, O., Ling, N., De, Souza, E. B. (1996). Urocortin interaction with corticotropin-releasing factor (CRF) binding protein: A novel mechanism for elevating "free" CRF levels in human brain. *Brain. Res.* **725,** 263-267.

Benedetto, C., Petraglia, F., Marozio, L., Chiarolini, L., Florio, P., Genezzani, A., Massobrio, M. (1994). Corticotropin-releasing hormone increases prostaglandin F2a activity on human myometrium *in vitro*. *Am. J. Obstet. Gynecol.* **171,** 126-131.

Benton, W. D., Davis, R. W. (1977). Screening lambda GT recombinant clones by hybridization to single plaques in situ. *Science.* **196,** 180-182.

Bird, A. P. (1986). CpG-rich island and the function of DNA methylation. *Nature.* **321,** 209-213.

Blackburn, M. B., Kingan, T. G., Bodnarw, W., Shabanowitz, J., Hunt, D. F., Kempe, T., Wagner, R. M., Raina, A. K., Schnee, Me, Ma. MC. (1991). Isolation and identification of a new diuretic peptide from the tobacco hornworm, *Manduca sexta*. *Biochem. Biophys. Res. Commun.* **181,** 927-932.

Blalock, J. E. (1989). A molecular basis for bidirectional communication between the immune and neuroendocrine system. *Physiol. Rev.* **69,** 1-32.

Bloomquist, B. T., Johnson, R. C., Mains, R. E. (1992). Rapid isolation of flanking genomic DNA using Biotin-RAGE, a variation of single-sided polymerase chain reaction. *DNA Cell Biol.* **11,** 791-797.

Brown, M. R., Fisher, L. A., Spiess, J., Rivier, J., Rivier, C., Vale, W. (1982). Comparison of the biologic actions of corticotrophin-releasing factor and sauvagin. *Regul. pept.* **4,** 107-114.

Buckler, A. J., Chang, D. D., Graw, S. L., Brook, D., Haber, D. A., Sharo, P. A., Housman, D. E. (1991). Exon amplification: A strategy to isolate mammalian genes based on RNA splicing. *Proc. Natl. Acad. Sci. USA.* **88,** 4005-4009.

Campbell, E. A., Linton, E. A., Wolfe, C. D. A., Scraggs, P. R., Jones, M. T., Lowry, R. J. (1987). Plasma corticotropin-releasing hormone concentrations during pregnancy and parturition. *J. Clin. Endocrinol. Metab.* **64,** 1054-1059.

Carey, M. (1998). Enhanceosome and transcriptional synergy. *Cell.* **92,** 5-8.

Chalmers, D. T., Lowenberg, T. W., De Soza, E. B. (1995). Localization of novel corticotropin-releasing factor receptor (CRF2) mRNA expression to specific subcortical nuclei in rat brain: comparison with CRF1 receptor mRNA expression. *J. Neurosci.* **15,** 6340-6350.

Chalmers, D. T., Lowenberg, T. W., Grigoriadis, D. E., Behan, D. P., De Souza, E. B. (1996). Corticotropin-releasing factor receptor: from molecular biology to drug design. *Trends in Pharmacol Sci.* **17,** 166-172.

Chandler, V. L., Maler, B. A., Yamamoto, K. R. (1983). DNA sequences bound specifically by glucocorticoid receptor in vitro render a heterologous promoter hormone responsive in vivo. *Cell.* **33,** 489-499.

Chang, C. P., Pearse, R.V., O'Connell, S., Rosenfeld, M.G. (1993). Identification of a seven trans-membrane helix receptor for corticotropin-releasing factor and sauvagine in mammalian brain. *Neuron.* **11,** 1187-1195.

Chatterjee, T K., Liu, X., Davisson, R. L., Fisher, R. A. (1997). Genomic organization of the rat pituitary adenylate cyclase-activating polypetide receptor gene. *J. Biol. Chem.* **272,** 12122-12131.

Chen, R., Lewis, K. A., Perrin, M. H., Vale, W. W. (1993). Expression cloning of a human corticotropin-releasing factor receptor. *Proc. Natl. Acad. Sci. USA.* **90,** 8967-8971.

Christy, R. J., Yang, V. W., Natambi, J. M., Geiman, D. E., Landschulz, W. H., Friedman, A. D., Nakabeppu, Y., Kelly, T. J., Lane, M. D. (1989). Differentiation-induced gene expression in 3T3-L1 preadipocytes: CCAAT/enhancer binding protein interacts with and activates the promoters of two adipocyte-specific genes. *Genes Dev.* **3,** 1323-1335.

Chu, H. M., Fischer, W. H., Osborne, T. F., Comb, M. J. (1991). NF-1 proteins from brain interact with the proenkephalin cAMP inducible enhancer. *Nucl. Acids Res.* **19,** 2721-2728.

Clifton, V. L., Read, M. A., Leitch, I. M., Boura, A. L. A., Robinson, P. J., Smith, R. (1994). Corticotropin-releasing hormone-induced vasodilatation in the human fetal placental circulation. *J. Clin. Endocrinol. Metabol.* **79,** 666-669.

Comb, M., Birnberg, N. C., Seasholtz, A., Herbert, E., Goodman, H. M. (1986). A cyclic AMP - and phorbol ester-inducible DNA element. *Nature.* **323,** 353-356.

Comb, M., Mermod, N., Hyman, S. E., Pearlberg, J., Ross, M. E., Goodman, H. M. (1988). Proteins bound at adjacent DNA elements act synerhistically to regulate human proenkephalin cAMP inducible transcription. *EMBO J.* **7,** 3793-3805.

Cummings, S., Elde, R., Ells, J., Lindvall, A. (1983). Corticotropin-releasing factor immunoreactivity is widely distributed within the central nervous system of the rat: an immunohistochemical study. *J. Neurosci.* **3,** 1355-1368.

Dautzenberg, F. M., Dietrich, K., Palchaudhuri, M. R., Spiess, J. (1997). Identification of two corticotropin-releasing factor receptors from *Xenopus laevis* with

high ligand selectivity: unusual pharmacology of the type 1 receptor. *J. Neurochem.* **69,** 1640-1649.

Davidson, E. H., Jacobs, H. T., and Britten, R, J. (1983). Very short repeats and coordinate induction of genes. *Nature*. **301,** 468-470.

de Groot, R. P., Auwerx, J., Karperien, M., Staels, B., Kruijer, W. (1991). Activation of junB by PKC and PKA signal transduction through a novel cis-acting element. *Nucl. Acids Res.* **19,** 775-781.

Deutsch, P. J., Hoeffler, J. P., Jameson, J. L., Habener, J. F. (1988). Cyclic AMP and phorbol ester-stimulated transcription mediated by similar DNA elements that bind distinct proteins. *Proc. Natl. Acad. Sci. USA*. **85,** 7922-7926.

Donaldson C. J., Sutton, S. W., Perrin, M. H., Corrigan, A., Lewis, K. A., Rivier, J., Vaughan, J. M., Vale, W. W. (1996). Cloning and charaterization of human urocortin. *Endocrinology*. **137,** 2167-2170.

Doucas, V., Spyrou, G., Yaniv, M. (1991). Unregulated expression of c-jun or c-fos proteins but not jun D inhibits oestrogen receptor activity in human breast cancer derived cells. *EMBO J.* **10,** 2237-2245.

De Souza, E. B. (1995). Corticotropin-releasing factor receptor: physiology, pharmacology, biochemistry and role in central nervous system and immune disorders. *Psychoneuroendocrinology*. **20,** 789-819.

Dumas, J. B., Edwards, M., Delort, J., Mallet, J. (1991). Oligodoxynucleotide ligation to sscDNAs: A new tool for cloning 5'ends of mRNA and for constructing cDNA libraries by in vitro amplification. *Nucleic. Acids Res.* **19,** 5227-5232.

Dunn, A. J., and Berridge, C. W. (1990). Physiological and behavioral responses to corticotrophin releasing factor administration: is CRF a mediator of anxiety of stress responses? *Brain Res. Rev.* **15,** 71-100.

Egerton, M., Needham, M., Evans, S., Millest, A., Cerillo, G., McPheat, J., Popplewell, M., Johnstone, D., Hollis, M. (1995). Identification of multiple human calcitonin receptor isoforms: Heterologous expression and pharmacological characterization. *J. Mol. Endocrinol.* **14,** 179-189.

Elsea, S. H and Patel, P. (1998). Human genome, chromosomes, and genes. *Principles of Molecular Medicine*, 1st edn, pp3-7. Edited by J.L. Jameson. Humana Press Inc., Totowa.

Feinberg, A. P. and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132,** 6.

Findlay, D. M., Martin, T. J., (1984). Relationship between internalization and calcitonin-induced receptor loss in T47D cells. *Endocrinology*. **115,** 72-83.

Fisch, T. M., Prywes, R., Simon, M. C., Roeder, R. G. (1989). Multiple sequence elements in the c-fos promoter mediate induction by cAMP. *Genes. Dev.* **3,** 198-211.

Fleischmann, R. D., Adams, M. D., Whit, O., Clayton, R. A., *et al.* (1995). Whole-genome random sequencing and assembly of haemophilus influenzaerd. *Science.* **269,** 496-512.

Foulkes, N. S., Borrelli, E. and Sassone-Corsi, P. (1991). CREM gene: use of alternative DNA binding domains generates multiple antagonists of cAMP-induced transcription. *Cell.* **64,** 739-749.

Frohman, M. A., Dush, M. K., and Martin, G. R. (1988). Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc. Nat.l Acad. Sci. USA.* **85,** 8998-9002.

Fuchs, A., Fuchs, F, Husslein, P., Soloff, M., Fernstrom, M. (1982). Oxytocin receptors and human parturition: a dual role for oxytocin in the initiation of labour. *Science*. **215,** 1396-1398.

Gorn, A. H., Rudolph, S. M., Flannery, M. R., Morton, C. C., Weremowicz, S.,

Wang, T. Z., Krane, S. M., Goldring, S. R. (1995). Expression of two human skeletal calcitonin receptor isoforms cloned from a giant cell tumor of bone. The first intracellular domain modulates ligand binding and signal transduction. *J. Clin. Invest.* **95,** 2680-2691.

Gornman, C. M., Moffat, L. F., Howard, B. H. (1982). Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. *Mol. Cell Biol.* **7,** 725-737.

Grammatopoulos, D., Milton, N. G. N., Hillhouse, E. W. (1994). The human myometrial CRH receptor: G Proteins and second messengers. *Mol. Cell. Endocrinol.* **99,** 245-250.

Grammatopoulos, D., Thompson, S., Hillhouse, E. W. (1995). The myometrium expresses multiple isoforms of the corticotropin-releasing hormone receptor. *J. Clin. Endocrinol. Metab.* **80,** 2388-2393.

Grammatopoulos, D., Stirrat, G. M., Williams, S. A., Hillhouse, E. W. (1996). The biological activity of the corticotropin-releasing hormone receptor-adenylate cyclase complex in human myometrium is reduced at the end of pregnancy. *J. Clin. Endocrinol. Metab.* **81,** 745-751.

Grammatopoulos, D., Dai, Y., Chen, J., Karteris, E., Papadopoulou, N., Easton, A. J., Hillhouse, E. W. (1998). Human CRH receptor: differences in subtype expression between pregnant and non-pregnant myometrium. *J. Clin. Endocrinol. Metab.* **83,** 2539-2544.

Grammatopoulos, D., and Hillhouse, E. W. (1999). Activation of protein kinase C by oxytocin inhibits the biological activity of the human myometrial corticotropin-releasing hormone receptor at term. *Endocrinology.* **140,** 585-594.

Grammatopoulos, D., Dai, Y., Randeva, H. S., Lenvine, M. A., Karteris, E., Easton, A. J., Hillhouse, E. W. (1999). A novel spliced variant of the type 1 corticotropin-releasing hormone receptor with a deletion in the seventh transmembrane domain present in the human pregnant term myometrium and fetal membranes. *Mol. Endocrinol.* **13,** 2189-2202.

Grigoriadis, D. E., and De Souza, E. B. (1989). Heterogeneity between brain and pituitary corticotropin-releasing factor receptor is due to differential glycosylation. *Endocrinology.* **125,** 1877-1888.

Guillemin, R., and Rosenberg, B. (1955). Humoral hypothalamic control of anterior pituitary: a study with combined tissue cultures. *Endocrinology.* **57,**599-607.

Hai, T., Curran, T. (1991). Cross-family dimerization of transcription factors Fos/Jun and ATF/CRE alters DNA binding specificity. *Proc. Natl. Acad. Sci.USA.* **88,** 3720-3724.

Harris, G. W. (1948). Neural control of the pituitary gland. *Physiol. Rev.* **28,** 139-179.

Harrison, S. C. (1991). A structural taxonomy of DNA binding domains. *Nature.* **353,** 715-719.

Harvey, R., and Darlison, M. G. (1991). Random-primed cDNA synthesis facilitates the isolation of multiple 5'-cDNA ends by RACE. *Nucleic. Acids Res.* **19,** 4002.

Hashimoto, K., Koga, M., Motomura, T., Kasayama, S., Kouhara, H., Ohnishi, T., Arita, N., Hayakawa, T., Sato, B., Kishimoto, T. (1995). Identification of alternatively spliced messenger ribonucleic acid encoding truncated growth hormone-releasing hormone receptor in human pituitary adenomas. *J. Clin. Endocrinol. Metab.* **80,** 2933-2939.

Hillhouse, E. W., Grammatopoulos, D., Milton, N. G. N., Quartero, H. W. P. (1993). The identification of a human myometrial corticotropin-releasing hormone

receptor that increases in a affinity during pregnancy. *J. Clin. Endocrinol. Metab.* **76,** 736-741.

Hjorth, A. L., Pears, C., Williams, J. G., Firtel, R. A. (1990). A developmentally regulated trans-acting factor recognizes dissimilar GC-rich elements controlling a class of cAMP-inducible *Dictyostelium* genes. *Genes Dev.* **4,** 419-432.

Ho, P., Fong, R. S. M., Kai, H. S. T., Lau, E. H. Y., Ngan, E. S. W., Cotton, C. U., Chow, B. K. C. (1999). The human secretin receptor gene: genomic organization and promoter characterization. *FEBS. Lett.* **455,** 209-214.

Hoeffler, J. P., Meyer, T. E., Yun, Y., Jameson, J. L., Habener, J. F. (1988). Cyclic AMP-responsive DNA -binding protein: structure based on a cloned placental cDNA. *Science.* **242,** 1430-1433.

Hunter, T. and Karin, M. (1992). The regulation of transcription by phosphorylation. *Cell.* **70,** 375-387.

Hurst, H. C. (1996). bZip proteins. *Protein Profile.* **3,** 1-72.

Imagawa, M., Chiu, R., Karin, M. (1987). Transcription factor AP-2 mediates induction by two different signal-transduction pathways: protein kinase C and cAMP. *Cell.* **51,** 251-260.

Inoue, H., Watana, N., Higashi, Y., FujiI-Kuriyama, Y. (1991). Structures of regulatory regions in the human cytochrome p-450scc (desmolase) gene. *Eur. J. Biochem.* **195,** 563-569.

Ishihara, T., Nakamura, S., Kaziro, Y., Takahashi, T., Takahashi, K., Nagata, S. (1991). Molecular cloning and expression of a cDNA encoding the secretin receptor. *EMBO J.* **10,** 1635-1641.

Ja Lowe, L. W., Pestell, R. G., Madison, L. D., Jameson, J, L. (1998). Mechanisms of hormone action. *Principles of Molecular Medicine,* 1st edn, pp419-431. Edited by J. L. Jameson. Humana press Inc. Totowa, NJ.

Jones, S. A., Challis, J. R. G. (1989). Local stimulation of prostaglandin production by corticotropin-releasing hormone in human fetal membranes and placenta. *Biochem. Biophys. Res. Commun.* **159,** 192-199.

Juppner, H., Abou-Samra, A, B., Freeman, M., Kong, X, F., Schipani, E., Richards, J., Kolakowski, Jr, L, F., Hock, J., Potts, Jr, J, T., Kronenberg, H, M., (1991). A G protein-linked receptor for parathyroid hormone and parathyroid hormone-related peptide. *Science.* **254,** 1024-1026.

Kagawa, N., Waterman, M. R. (1990). cAMP-dependent transcription of the human CYP21B (P-450C21) gene requires a cis-regulatory element distinct from the consensus cAMP-regulatory element. *J.Biol. Chem.* **265,** 11299-11305.

Karim, S., Hillier, K., Trussell, R., Patel, R., Tamusange, S. (1970). Induction of labour with prostaglandin E_2. *J. Obstet. Gynaecol. Br. Commonw.* **77,** 200-210.

Karteris, E., Grammatopoulos, D., Dai, Y., Olah, K. B., Ghobara, T. B., Easton, A. J., Hillhouse, E. W. (1998). The human placenta and fetal membranes express the corticotropin-releasing hormone receptor 1α (CRH-R1α) and the CRH-C variant receptor. *J. Clin. Endocrinol. Metab.* **83,** 1376-1379.

Kasai, H., Isono, S., Kitakawa, M., Mineno, J., Akiyama, H., Kurnit, D. M., Berg, D. E., Isono, K. (1992). Efficient large-scale sequencing of the escherichia coli genome: implementation of transponson-and PCR-based strategy for the analysis of order λ phage clones. *Nucl. Acids Res.* **20,** 6509-6515..

Kataoka, H., Troetshler, R. G., Li, T. P., Kramer, S. J., Corney, R. L., Schooley, D. A. (1989). Isolation and identification of a diuretic hormone from the tobacco hornworm, *Manduca sexta. Proc. Natl. Acad. Sci. USA.* **86,** 2976-2980.

Kaziro, Y., Itoch, H., Kozasa, T., Nakafuku, M., Satoh, T. (1991). Structure and function of signal transducing GTP-binding proteins. *Annu. Rev.Biochem.* **60,** 349-400.

Kilpatrick, G. J., Dautzenberg, F. M., Martin, G. R., and Eglen R. M. (1999). 7TM receptors: the splicing on the cake. *Trends Pharmacol. Sic.* **20,** 294-301.

Kishimoto, T., Pearse, R. V., Lin, C. R., Rosenfeld, M. G. (1995). A sauvagine/corticotropin-releasing factor receptor expressed in heart and skeletal muscle. *Proc. Natl. Acad. Sci. USA.* **92,** 1108-1112.

Kong, X. F., Schipani, E., Lanske, B., Joun, H., Karperien, M., Defize, L. H., Juppner, H., Potts, Jr., L. T., Segre, G.V., Kronenberg, H. M., Abou-Samura, A. B. (1994). The rat, mouse and human genes encoding the receptor for parathyroid hormone and parathyroid hormone-related peptide are highly homologous. *Biochem. Biophys. Res. Commun.* **200,** 1290-1299.

Kornberg, T. B. (1993). Understanding the homeodomain. *J. Biol. Chem.* **268,** 26813-26816.

Kostich, W. A., Chen, A., Sperle, K., Largent, B. L. (1998). Molecular identification and analysis of a novel human corticotropin-releasing factor receptor: The CRF2γ receptor. *Mol. Endocrinol.* **12,** 1077-1085.

Kuestner, R. E., Elrod, R. D., Grant, F. J., Hagen, F. S., Kuijper, J. L., Matthewes, S. L., O'Hara, P. J., Sheppard, P. O., Stroop, S. D., Thompson, D. L., Whitmore, T. E., Findly, D. M., Houssami, S., Sexton, P. M., Moore, E. E. (1994). Cloning and characterization of an abundant subtype of the human calcitonin receptor. *Mol. Pharmacol.* **46,** 246-255.

Kurten, R. C., Levy, L. O., Dshey, J., Durica, J. M., Richards, J. S. (1992). Identification and characterization of the GC-rich and cyclic adenosin 3'5'-monophosphate (cAMP)-inducible promoter of the type IIB cAMP-dependent protein kinase regulatory subunit gene. *Mol. Endocrinol.* 537-549.

Lania, L., Majello, B., and Luca, P. D. (1997). Transcriptional regulation by the Sp family proteins. *Int. J. Biochem. Cell Biol.* **29,** 1313-1323.

Latchman, D. S. (1997). Transcription factors. *Int. Biochem. Cell Biol.* **12,** 1305-1312.

Lederis, K., Letter, A., McMaster, D., Moore, G. (1982). Complete amino acid sequence of urotensin-1, a hypotensive and corticotronpin releasing neuropeptide from *Catostomus. Science.* **218,** 162-164.

Lee, K. A., Hai, T. Y., SivaRaman, L. (1987). A cellular protein, activating transcription factor, activates transcription of multiple E1A-inducible adenovirus early promoters. *Pro. Natl. Acad. Sci. USA.* **84,** 8355-8359.

Lenz, H. J., Fisher, L. A., Vale, W. W., Brown, M. R. (1985). Corticotropin-releasing factor, sauvagine and urotensin 1: effects on blood flow. *Am. J. Physiol.* **249,** R85-R90.

Liaw, C. W., Lovenberg, T. W., Barry, G., Oltersdorf, T., Grigoriadis, D. E., De Souza, E. B. (1996). Cloning and characterization of the human corticotropin-releasing factor-2 receptor complementary deoxyribonucleic acid. *Endocrinology.* **137,** 72-77.

Lin, Y. S., Green, M. R. (1988). Interaction of a common cellular transcription factor, ATF, with regulatory elements in both Elaand cyclic AMP-inducible promoters. *Pro. Natl. Acad. Sci. USA.* **85,** 3396-3400.

Linton, E. A., Perkins, A. V., Woods, R. J., Eben, F., Wolfe, C. D. A., Behan, D. P., Potter, E., Vale, W. W., Lowry, P. J. (1993). Corticotropin-releasing hormone binding protein (CRH-BP): plasma levels decrease during the third trimester of normal human pregnancy. *J. Clin. Endocrinol. Metab.* **76,** 260-262.

Littlewood, T., and Evan, G. (1995). Helix-loop-helix. *Protein profile.* **2,** 621-702.

Loh, E. Y., Elliott, J. F., Cwirla, S., Lanier, L. L., Davis, M.M. (1989). Polymerase chain reaction with single-side specificity: analysis of T cell receptor δ chain. *Science.* **243,** 217-220.

Lok, S., Kuijper, J. L., Jelin, L. J., Kramer, J. M., Whitmore, T. E., Sprecher, C. A., Mathewes, S., Grant, F. J., Biggs, S. H., Rosenberg, G. B., Sheppard, P. O., O'Hara, P. J., Foster, D. C., Kindsvogel, W. (1994). The human glucagon receptor encoding gene: structure, cDNA sequence and chromosomal localization. *Gene.* **140,** 203-209.

Lovenberg, T. W., Liaw, C. W., Grigoriadis, D. E., Clevenger, W., Chalmers, D. T., De Suza, E. B., Oltersdorf, T. (1995a). Cloning and characterization of a functionally distinct corticopin releasing factor receptor subtype from rat brain. *Proc. Natl Acad. Sci. USA.* **92,** 836-840.

Lovenberg, T. W., Chalmers, D. T., Liu, C., De Souza, E. B. (1995b). CRF2α and CRF2β receptor mRNAs are differentially distributed between the rat central nervous system and peripheral tissues. *Endocrinology.* **136,** 4139-4142.

Lowe, W. L., Pestell, R. G., Madison, L. D., Jameson, J. L. (1998). Mechanisms of hormone action. *Principles of Molecular Medicine.* 1st edn, pp. 419-431. Edited by J. L. Jameson. Humana Press Inc., Totowa, NJ.

Lund, J., Ahlgren, R., Wu, D. H., Kagimoto, M., Simpson, E. R., Waterman, M, R. (1990). Transcriptional regulation of the bovine CYP17 (P-450 (17) Alpha) gene. Identification of two cAMP regulatory regions lacks the consensus cAMP-responsive element (CRE). *J. Biol. Chem.* **265,** 3304-3312.

Lynch, J. P., Lala, D. S., Peluso, J. J., Luo, W., Parker, K. L., White, B. A. (1993). Steroidogenic factor 1, an orphan nuclear receptor, regulates the expression of the rat aromatase gene in gonadal tissues. *Mol. Endocrinol.* **7,** 776-786.

Macrae, A. D., Brenner, S. (1994). One armed PCR (OA-PCR): amplification of genomic DNA from a single primer domain. *Genomics.* **24,** 176-178.

Manen, D., Palmer, G., Bonjour, J. P., Rizzoli, R. (1998). Sequence and activity of parthyroid hormone/parathyroidhormone-related peptide receptor promoter region in the human osteoblast-like cells. *Gene.* **218,** 49-56.

Mayo, K. E. (1992). Molecular cloning and expression of a pituitary-specific receptor for growth hormone-releasing hormone. *Mol. Endocrinol.* **6,** 1734-1744.

McCuaig, K. A., Lee, H. S., Clarke, J. C., Assar, H., Horsford, J., White, J. H. (1995). Parathyroid hormone/parathyroidhormone-related peptide receptor gene transcripts are expressed from tissue-specific and ubiquitous promoters. *Nucl. Acid Res.* **23,** 1948-1955.

McKnight, S. G., Clegg, C. H., Uhler, M. D., Chrivia, J. C., Cadd, G. G. (1988). Analysis of the cAMP dependent protein kinase system using molecular genetic approaches. *Rec. Progr. Horm. Res.* **44,** 307-335.

McLean, M., Thompson, D., Zhang, H. P., Brinsmead, M., Smith, R., (1994). Corticotropin-releasing hormone and beta-endorphin in labour. *Eur. J. Endocrinol.* **131,** 167-172.

McLean, M., Bisits, A., Davies, J., Woods, R., Lowry, P., Smith, R. (1995). A placental clock controlling the length of human pregnancy. *Nature Medicine.* **1,** 460-463.

Meyer, A. H., Ullmer, C., Schmuck, K., Morel, C., Wishart, W., Lubbert, H., Engels, P. (1997). Localization of the human CRF2 receptor to 7p21-p15 by radiation hybrid mapping and FISH analysis. *Genomics.* **40,** 189-190.

Mezei, L. M. and Storts, D. R. (1994). Cloning PCR products. In: PCR Technology: Current Innovation. 1st eds, pp. 21-27. Edited by Griffin, H. G. and Griffin, A. M. CRC Press, Boca Raton, FL.

Mitchell, P. J., Wang, C., Tjian, R., (1987). Positive and negative regulation of transcription in vitro: enhancer binding protein AP-2 is inhibited by SV40 T antigen. *Cell.* **50,** 847-861.

Miyata, I., Shiota, C., Ikeda, Y., Oshida, Y., Chaki, S., Okuyama, S., Inagami, T. (1999). Cloning and characterization of a short variant of the corticotronpin-releasing factor receptor subtype from rat amygdala. *Biochem. Biophys. Res. Commun.* **256,** 692-696.

Montecucchi, P. C., Heushen, A. (1981). Amino acid composition and sequence analysis of sauvagion, a new active peptide from the skin of *Phyllomedusa sauvagei.* *Int. J. peptide protein Res.* **18,** 113-120.

Montminy, M. R., Sevarino, K. A., Wagner, J. A., Mandel, G., Goodman, R. H. (1986). Identification of a cyclic-AMP-responsive element within the rat somatostatin gene. *Proc. Natl. Acad. Sci. USA.* **83,** 6682-6686.

Moore, E. E., Kuestner, R. E., Stroop, S.D., Grant, F. J., Matthewes, S. L., Brady, C. L., Sexton, P. M., Findlay, D. M. (1995). Functionally different isoforms of the human calcitonin receptor result from alternative splicing of the gene transcript. *Mol. Endocrinlo.* **9,** 959-968.

Mount, S. M. (1982). A catalogue of splice junction sequences. *Nucl. Acids Res.* **10,** 459-473.

Mueller, P. R., Wold, B. (1991). Ligation-mediated PCR: applications to genomic footprinting. *Methods.* **2,** 20-31.

Muller, M. M., Gerster, T. and Schaffner, W. (1988). Enhancer sequences and the regulation of gene transcription. *Euro. J. Biochem.* **176,** 441-446.

Murray, V. (1989). Improved double stranded sequencing using the linear polymerase chain reaction. *Nucl. Acids Res.* **17:** 88-89.

Nakamura, M., Hashimoto, T., Nakajima, T., Ichii, S., Furuyama, J., Ishihara, Y., Kakudo, K. (1995). A new type of human calcitonin receptor isoform generated by alternative splicing. *Biochem. Biophys. Res. Commun.* **209,** 744-751.

Nappi, R. E., and Rivest, S. (1995). Stress-induced genetic expression of a selective corticotropin-releasing factor receptor subtype within the rat ovaries: An effect dependent on the ovulatory cycle. *Biol. Reprod.* **53,** 1417-1428.

Novina, C. D., and Roy, A. L. (1996). Core promoters and transcriptional control. *Trends Genet.* **12,** 351-355.

Ochman, H., Gerber, A. S., Hartl, D. L. (1988). Genetic applications of an inverse polymerase chain reaction. *Genetics.* **120,** 621-623.

Ohara, O., Dorit, R, L., and Gilbert, W. (1989). One-sided polymerase chain reaction: the amplification of cDNA. *Proc. Natl Acad. Sci. USA.* **86,** 5673-5677.

Okamoto, E., Takagi, T., Makino, T. (1989). Immunoreactive corticotropin-releasing hormone, adrenocorticotropin and cortisol in human plasma during pregnancy and delivery and postpartum. *Horm. Metab. Res.* **21,** 566-572.

Okawara, Y., Morley, S. D., Turzio, L. O., Zwiers, H., Lederis, K., Richter, D. (1988). Cloning and sequence analysis of cDNA for corticotropin releasing factor precursor from the teleost fish *Catastomus commersoni. Proc. Natl. Acad. Sci. USA.* **85, 8439-8443.**

Olschowka, J. A., O'Donohue, T. L., Mueller, G. P., Jacobowitz, D. M. (1982). The distribution of corticotropin-releasing factor-like immunoreactive neurons in rat brain. *Peptides*. **3,** 995-1015.

Orth, D. N. (1992). Corticotropin-releasing hormone in humans. *Endocr. Rev.* **13,** 164-191.

O'Shea-Greenfield, A., and Smale, S. T. (1992). Roles of TATA and initiator elements in determining the start site location and direction of RNA polymerase II transcription. *J.Biol. Chem,* **267,** 1391-1402.

Owens, M. J., Mulchahey, J. J., Kasckow, J. W., Plotsky, P. M., Nemeroff, C. B. (1995). Exposure to an antisense oligonucleotide decreases corticotropin-releasing factor receptor binding in rat pituitary cultures. *J. Neurochem.* **64,** 2358-2361.

Parker, M. G. (1993). Steroid and related receptors. *Curr. Opin. Cell. Biol.* **1,** 512-518.

Pei, L. and Melmed, S. (1995). Characterization of the rat vasoactive intestinal polypeptide receptor gene 5'region. *Biochem. J.* **308,** 719-723.

Perkins, A. V., and Linton, E. A. (1995). Placental corticotropin-releasing hormone: there by accident or design? *Journal of Endocrinolory*. **147,** 377-381.

Perrin, M. H., Haas, Y., Rivier, J., Vale, W. W. (1986). Corticotropin-releasing factor binding to the anterior pituitary receptor is modulated by divalent cations and guanyl nucleotides. *Endocrinology*. **118,** 1171-1179.

Perrin, M. H., Donaldson, C. J., Chen, R., Lewis, K. A., Vale, W. W. (1993). Cloning and functional expression of a rat brain corticotropin-releasing factor (CRF) receptor. *Endocrinology*. **133,** 3058-3061.

Perrin, M. H., Donaldson, C. J., Chen, R., Blount, A., Berggren, T., Bilezikjian, L., Sawchenko, P. E., Vala, W. W. (1995). Identification of a second corticotropin-releasing factor receptor gene and characterization of a cDNA expressed in heart. *Proc. Natl Acad. Sci. USA*. **92,** 2969-2973.

Petersenn, S., Rasch, A. C., Heyens, M., Schulte, H. M. (1998). Structure and regulation of the human growth hormone-releasing hormone receptor gene. *Mol. Endocrinol.* **12,** 233-247.

Petraglia, F., Sawchenko, P., Rivier, J., Vale, W. (1987). Evidence for local stimulation of ACTH secretion by corticotropin-releasing factor in human placenta. *Nature*. **328,** 717-719.

Pisegna, J. R., Wank, S. A., (1996). Cloning and characterization of the signal transduction of four splice variants of the human pituitary adenylate cyclase activating polypeptide receptor. Evidence for dual coupling to adenylate cyclase and phospholipase C. *J. Biol. Chem.* **271,** 17267-17274.

Potter, E., Behan, D. P., Fischer, W. H., Linton, E. A., Lowry, P. J., Vale, W.W. (1991). Cloning and characterization of the cDNA for human and rat corticotropin-releasing factor-binding proteins. *Nature*. **349,** 426-432.

Potter, E., Sutton, S., Donaldson, C., Chen, R., Perrin, M., Lewis, K., Sawchenko, P. E., Vale, W. W. (1994). Distribution of corticotropin-releasing factor receptor mRNA expression in the rat brain and pituitary. *Proc. Natl Acad. Sci. USA*. **91,** 8777-8781.

Probst, W. C., Snyder, L. A., Schuster. D. I., Brosius. J., Sealfon. S. C. (1992). Sequence alignment of the G-protein coupled receptor superfamily. *DNA Cell Biol.* **11,** 1-20.

Pugh, B. F. and Tjian. R. (1990). Mechanism of transcriptional activation by Sp1: evidence for coactivators. *Cell.* **61,** 1187-1197.

Quandt, K., Frech, K., Wingender, E., Werner, T. (1995). MATIND and MATINSPECTOR- new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic. Acids Res.* **23,** 4878-4884.

Quartero, H., Srivatsa, G., Gillham, B. (1992). Role for cyclic adenosine monophosphate in the synergistic interaction between oxytocin and corticotropin-releasing factor in isolated human gestational myometrium. *Clin. Endocrinol (Oxf).* **36,** 141-145.

Radler-Pohl, A., Gebel, S., Sachsenmaier, C., Konig, H., Kramer, M., Oehler, T., Streile, M., Ponta, H., Rapp, U., Rahmsdorf, H. J., Cato, A. C. B., Angel, P., Herrlich, P. (1993). The activation and activity control of AP-1 (Fos/Jun). *Ann. NY. Acad. Sci.* **684,** 127-148.

Rees, L. H., Burke, C. W., Chard, T., Evans, S. W., Letchworth, A. T. (1975). Possible placental origin of ACTH in normal human pregnancy. *Nature*. **1975,** 620-622.

Rehfuss, R. P., Walton, K. M., Loriaux, M. M., Goodman, R. H. (1991). The cAMP-regulated enhancer-binding protein ATF-1 activates transcription in response to cAMP-dependent protein kinase A. *J. Biol. Chem.* **266,** 18431-18434.

Rivier, C., Rivier, J., Lederis, K., Vale, W. W. (1983). *In vitro* and *in vivo* ACTH-releasing activity of ovine CRF, sauvagine and urotensin-1. *Regul. Pept.* **5,** 139-143.

Roesler, W. J., Park, E. A. (1998). Hormone response units: one plus one equals more than two. *Mol. Cell Biochem.* **178,** 1-8.

Rose, J. K., Buonocore, L., and Whitt, M. M. (1991). A new cationic liposome reagent mediating nearly quantitative transfection of animal cells. *Biotechniques.* **10,** 520-525.

Rosenthal, W., Antaramian, A., Gilbert, S., Birnbaumer, M. (1993). Nephrogenic diabetes insipidus. A vasopressin receptor unable to stimulate adenylyl cyclase. *J. Biol. Chem.* **268,** 13030-13033.

Ross, P. C., Kostas, C. M., Ramabhadran, T. V. (1994). A variant of the human corticotropin-releasing factor (CRF) receptor: cloning, expression and pharmacology. *Biochem. Biophys. Res. Commun.* **205,** 1836-1842.

Sadovsky, Y., Crawford, P. A., Woodson, K. G., Polish, J. A., Clements, M. A., Tourtellotte, L. M., simburger, K., Milbrandt, J. (1995). Mice deficient in the orphan receptor steroidogenic factor 1 lack adrenal glands and gonads but express P450 side-chain-cleavage enzyme in the placenta and have normal embryonic serum levels of corticosteroids. *Proc. Natl. Acad. Sci. USA.* **92,** 10939-10943.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science.* **239,** 487-491.

Sakai, K., Yamada, M., Horiba, N., Wakui, M., Demura, H., Suda, T. (1998). The genomic organization of the human corticopin-releasing factor type-1 receptor. *Gene.* **219,** 125-130.

Samama, P., Cotecchia, S., Costa, T., Lefkowitz, R. J. (1993). A mutation-induced activated state of the β2-adrenergic receptor. Extending the ternary complex model. *J. Biol. Chem.* **268,** 4625-4636.

Sambrook, J., Fritsch, E. F., Maniatis, T. (1989). Molecular Cloning: A Laboratory manual. 2nd edn, Edted by Sambrook, J., Fritsch, E. F., Maniatis, T. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sanger, F., Niklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Pro. Natl. Acad. Sci. USA.* **74,** 5463-5467.

Sassone-Corsi, P. (1995). Transcription factors responsive to cAMP. *Ann. Rev. Cell. Dev. Biol.* **11,** 355-357.

Sawchenko, P. E., Arias, C., Van Pett, K., Viau, V., Chen, R., Donaldson, C., Blount, A., Bilezikjian, L., Perrin, M., Vale, W. W. (1995). Distribution of mRNAs encoding two distints CRF receptors in brain and pituitary of rat and mouse. *77th Annual Meeting of the Endocrine Society. Washington.* Abstract: P1-558.

Segre, G. V., Goldring, S. R. (1993). Receotors for secretin, calcitonin, parathyroid hormone (PTH)/PTH-related peptide, vasoactive intestinal peptide, glucagon-like peptide1, growth hormone releasing hormone, and glucagon belong to the newly discovered G-protein likeed receptor family, *Trends Endocrinol Metab.* **4,** 309-314.

Schaefer, B. C. (1995). Revolutions in rapid amplification of cDNA ends: new strategies for polymerase chain reaction cloning of full-length cDNA end. *Anal. Biochem.* **227,** 255-273.

Sehgal, A., Patil, N., and Chao, M. (1988). A constitutive promoter directs expression of the nerve growth factor receptor gene. *Mol. Cell Biol.* **8,** 3160-3167.

Short, J. M., Wynshaw, B. A., Short, H. P., Hanson, R. W. (1986). Characterization of the phosphoenolpyruvate carboxykinase (GTP) promoter-regulatory region. II. Identification of cAMP and glucocorticoid regulatory domains. *J. Biol. Chem.* **261,** 9721-9726.

Shyu, J. F., Inoue, D., Baron, R., Horme, W. C. (1996). The deletion of 14 amino acids in the seventh transmembrane domain of a naturally occurring calcitionin receptor isoform alters ligand binding and selectively abolishes coupling to phospholipase C. *J. Biol. Chem.* **271,** 31127-31134.

Silver, B. J., Boker, J. A., Virgin, J. B., Vallen, E. A., Milsted, A., Nilson, J. H. (1987). Cyclic AMP regulation of the human glycoprotein hormone alpha-subunit gene is mediated by an 18-base-pair element. *Pro. Natl. Acad. Sci. USA.* **84,** 2198-2202.

Slominski, A., Ermak, G., Hwang, J., Chakraborty, A., Mazurkiewicz, J. E., Mihm, M. (1995). Proopioomelanocortin, corticotropin-releasing hormone and corticotropin releasing hormone receptor genes are expressed in human skin. *FEBS Lett.* **374,** 113-116.

Smale, S. T., Baltimore, D. (1989). The "initiator" as a transcription control element. *Cell.* **57,** 103-113.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98,** 503-517.

Spengler, D., Waeber, C., Pantaloni, C., Holsboer, F., Bockaert, J., Seeburg, P. H., Journot, L. (1993). Differential signal transduction by five splice variants of the PACAP receptor. *Nature.* **365,** 170-175.

Spiegel, A. M., Shenker, A., Weinstein, K. S., (1992). Receptor effector coupling by G proteins: implications for normal and abnormal signal transduction. *Endocr. Rev.* **13,** 536-565.

Sreedharan, S. P., Huang, Jx., Cheung, M. C., Goetzl, E. J. (1995). Structure, expression, and chromosomal localization of the type I human vasoactive intestinal peptide receptor gene. *Proc. Natl Acad. Sci. USA.* **92,** 2939-2943.

Stenzel, P., Kesterson, R., Yeung, W., Cone, R. D., Rittenberg, M. B., Stenzel-Poore, M. P. (1995). Identification of a novel murine receptor for corticotropin releasing hormone expressed in heart. *Mol. Endocrinol.* **9,** 637-645.

Stevens, M. Y., Challis, J. R. G., Lye, S. J. (1998). Cortitropin-releasing hormone receptor subtype 1 is significantly up-regulated at the time of labour in the human myometrium. *J. Clin. Endocrinol. Metab.* **83,** 4107-4115.

Sugawara, T., Kiriakidou, M., McAllister, J. M., Kallen, C. B., Strauss, J. F. (1997). Multiple steroidogenic factor 1 binding elements in the human steroidogenic acute regulatory protein gene 5'-flanking region are required for maximal promoter activity and cyclic AMP responsiveness. *Biochemistry.* **36,** 7249-7255.

Sutton, S. W., Behan, D. P., Sabine, L. L., Kaiser, R., Corrigan, A., Lowry, P., Potter, E., Perrin, M. H., Rivier, J., Vale. W. (1995). Ligand requirements of the human corticotropin releasing factor-binding protein. *Endocrinology.* **136,** 1097-1102.

Tsai-Morris, C. H., Buczko, E., Geng, Y., Gamboa-Pinto, A., Dufau, M. L. (1996). The genomic structure of the rat corticotropin-releasing factor receptor. A member of the class II G proptein -coupled receptors. *J. Biol. Chem.* **271,** 14519-14525.

Tsukada, T., Fink, J. S., Mandel, G., Goodman, R. H. (1987). Identification of a region in the human vasoactive intestinal polypeptide gene responsible for regulation by cyclic AMP. *J. Biol Chem.* **262,** 8743-8747.

Turnbull, A. V. and Rivier, C. (1997). Corticotrophin-releasing factor and endocrine responses to stress: CEF receptors, binding protein, and related peptides. *Experimental Biology and Medicine.* **215,** 1-10.

Valdenaire, O., Giller, T., Breu, V., Gottowik, J., Kilpatric, G. (1997). A new functional isoforms of the human CRF2 receptor for corticotropin-releasing factor. *Biochem. Biophys. Acta.* **1352,** 129-132.

Vale, W., Spiess, J., Rivier, C., Rivier, J. (1981). Characterization of a 41-residue ovine hypothalamic peptide that stimulates secretion of corticotropin and β-endorphin. *Science.* **213,** 1394-1397.

Vale, W., Rivier, C., Brown, M. R., Spiess, J., Koob, G., Swanson, L., Bilezikjian, L., Bloom, F., Rivier, J. (1983). Chemical and biological characterization of corticotropin- releasing factor. *Recent Prog. Horm. Res.* **39,** 245-270.

Vamvakopoulos, N. C., Sioutopoulou, T. O. (1994). Human corticotropin-releasing hormone receptor gene (CRHR) is located on the long arm of chromosome 17(17q12-qter). *Chromosome Res.***2,** 471-471.

Vaughan, J., Donaldson, C., Bittencourt, J., Perrin, M. H., Lewis, K., Sutton, S., Chen, R., Turnbull, A. V., Lovejoy, D., Rivier, C., Rivier, J., Sawchenko, P. E., Vale, W. (1995). Urocortin, a mammalian neuro-peptide related to fish urotensin I and to corticotropin- releasing factor. *Nature.* **378,** 287-292.

Vita, N., Laurent, P., Lefort, S., Chalon, P., Lelias, J-M., Kaghad, M., Le Fur, G., Caput, D., Ferrara, P. (1993).Primary structure and functional expression of mouse pituitary and human brain corticotropin-releasing factor receptors. *FEBS Lett.* **335,**1-5.

Weis, L., and Reinberg, D. (1992). Transcription by RNA polymerase II: initiator-directed formation of transcription-competent complexes. *FASEB J.* **6,** 3300-3309.

Wikland, M., Lindblom, B., Wilhelmsson, L., Wiqvist, N. (1982). Oxytocin, prostaglandins, and contractility of the human uterus at term pregnancy. *Act Obstet Gynaecol Scandinav.* **61,** 467-472.

Williams, T., Admon, A., Luscher, B., Tjian, R. (1988). Cloning and expression of AP-2, a cell-type-specific transcription factor that activates inducible enhancer elements. *Genes. Dev.* **2,** 1557-1569.

Williams, T., Tjian, R. (1991). Analysis of the DNA-binding and activation properties of the human transcription factor AP-2. *Genes. Dev.* **5,** 670-682.

Xiong, Y., Xie, L. Y., Abou-Samra, A. (1995). Signalling properties of mouse and human corticotropin-releasing factor (CRF) receptors: decreased coupling efficiency of human type II CRF receptor. *Endocrinology.* **136,** 1828-1834.

Yu, J., Xie, L. Y., Abou-Samra, A. (1996). Molecular cloning of a type A chicken corticotropin-releasing factor receptor with high affinity for urotensin I. *Endocrinology.* **137,** 192-197.

Zhang, P. and Mellon, S. H. (1996). The orphan nuclear receptor steroidogenic factor-1 regulates the cyclic adenosine 3',5'-monophosphate-mediated transcriptional activation of rat cytochrome P450c17 (17 alpha-hydroxylase/c17-20 lyase). *Mol. Eodocrinol.* **10,** 147-158.

Zinn, K., DiMaio, D. and Maniatis, T. (1983). Identification of two distinct regulatory regions adjacent to the human β-interferon gene. *Cell.* **34,** 865-879.

Zolinierowicz, S., Cron, P., Solinas, T. S., Fries, R., Lin, H. Y., Hemmings, B. A. (1994). Isolation, characterization, and chromosomal localization of the porcine calcitonin receptor gene. Identification of two variants of the receptor generated by alternative splicing. *J. Biol. Chem.* **269,** 19530-19538.

Appendix DNA sequence of the 5'end of the CRH-R2 gene

1 tccagtccct aacccccagc cactactggc atgaggggtc cctcagggcc cccaggcctc
61 ctctacgtcc cacacctcct cctctgcctg ctctgcctcc tcccaccgcc gctccaatac
121 gtaagtgcag ggagtgggag tccacgaggt ggggcgccag gagcacggag ctcccagctc
181 tcatgactag gggcgggggt ctcagggcct ggtggctggg ctggatcggg gatccttgtg
241 tctgatgatg aattggagat cccgtgcctg ggtgaggggg tctcagggcc tggtggctgg
301 gctggatcgg ggatccttgt gtctgatgat gaattggaga tcccgtgcct gggtgagggg
361 gcaggggaag ggaagatgaa actgtgaggc tggccaggat gagactggga gacccctgtg
421 gataagcaga gggttctggg gtgacatcct agcatccaga acttgactac ctcagggtcc
481 aagattccct gtctcctatc cattctctcc ccacccccat gacccccctc catctctcct
541 gcaagatcct tgacttcttt ccccagccca cccctagttc tgacccttac ccccagggac
601 ccgcagagag ccacctcctg cccctgctgc cttgtctcca tccttcctct cagtgcctgg
661 aatcaggcag cccccagtac cagatgagcg atggggacag agtgggctct gggagaacca
721 gcctctccaa gggtttctca agactcagag cttcatccag ccacaccccg tgcctttgtc
781 cagagtcagg ggccaccgat ggaaggtggg gagaagggac acagagcgac catcaaaaga
841 gagctgaaga agtaggggag aggctggtgg gagtggtcca cctccagcct tctccccatt
901 ttacagatgg ggaaactgag gctcagagtg gtagggtgac ctgtcctagg cctcacaatc
961 tggaaagtag tgcctgtgga agctaagagg cttgtcgctg tcctgactct gccacagacc
1021 agctttgggg accattctct ggggcttagt gtccctatac gtataaggag gataatacat
1081 tcaacagtaa ttcaacagat aattgcctac aaaaccaatc cagatccctg ctttgctgct
1141 taatgtgtga ccttggacag gttactcaaa tcctcagttt cctcatcagt aaaatgagga
1201 tcacaacggt gcttactgaa caagcttgtt gtgaggaaca ttgaggtaat acaagtaaag
1261 caagcagcat catgcttggt gtggtgtgag cacgcagtga atgccagcta tgcttttgtg
1321 ccagccgcca cagatacagc atgaggacat gagggtccct gccttcctgg gggcctggcc
1381 aggtggagga gagaggtggg aggcaggtag tatgctgagc tgtggtcagt gtccgatagg
1441 cagatggccc ccagacctgc tgagtccttt gagtgacttt gcagactgca attgggatgg
1501 gtgaaggttg tacaggggggc tcaccatcca cccctctgcc cccaagacat cttgggcttt
1561 gagctctctt gccctctccc acttgttttc cagataggtg ccttccagga gatccaagtg
1621 ttggggcaag ggttgcccct cagcccgtcc ctctggggtc catgctgcca gccgcccaag
1681 ctggttagac ggaggaagct tctgctggct gctccagagg tgccctagga ttctcagggc
1741 aacactgtgg tctcagcgtc ccttgaggtc atcccaggcc ctgcgcctag ctctcctgcg
1801 gatcccggag tgggatagag accagttact tacccacagc cactcagcat gtaagcagct
1861 ggagagcact taaacctcgg cctttctgac gccaaagccc agccttcctc tcacggggcc
1921 ttcatggcaa tgggactttt tcccgggtcc tgagacgatg agacctggag caggcagtcg
1981 ggagggaagg cagcccagga gagcccagcc ttggaggctt ctcagttatt catccagtgt
2041 ttaaggagga ctgcccatgg ggaggatcca gggtctgaac gctcatctgg ttcctgttct
2101 tgtggtgttt atagcccaga agcgggtgtg gcaggaaaga aaggagtctg tgtttacaag
2161 ttccagtaaa tgggctcaga agaagcatgt gtaggtgtga gggcccttga caggatgggg
2221 gagcctggtg ggaggtgggc agcatagcag agttggagat gggtagtcca agggaaggga
2281 taggggagca ggttgggggt aaggccccag ggccccaacc ctgaaaagac attatgtggg
2341 agactgcaga gctgtatggg tcctggtggg atctgaggac atggggagct ctctgagcag
2401 gggaggtgcc tggtctgatc tgcaattgta agggttccag ctggtgcttt tgagtgaaga
2461 ggatgccctg gcccagggca cactgagagc ctggaggaac ccacccagcc ttgaccctgg
2521 tgctccctat tccccgtgga ggccttcggg aagctggtcc ggaggcaggt ggtaatctag
2581 cagctgttaa gtggtgttag gcgccttact caaacaccgg gaaggtggga ccacttattt
2641 acagacagga aacaagttta aagaggctca gcaatgccac ttgcctccca ctctgtgaag
2701 atcacaggag cctctcagcc cctgctgact aagtccagct ccttggagct ttggacagcg

2761 agacaccccc cagggtctca ggatggtgcc tctgtgacct cttgccccag gtcctggctt
2821 ctggtgggcc catctgcccc aagaaggctc aacaaagaga acagaacaca gagaagtccc
2881 tggcagtccc tggaccctcc ctgacctgtc cggctcacac caaagaccga gtccacagtg
2941 tcccctaaca tacacacact gagggtctgc agagccgggg ctccagggcc atttgccagc
3001 agctaatgca ttgaatgact tttcccaaaa attcaactcc cttctgagcg ttttctagcc
3061 tcactagcta atttagtctc tagttttgcc attttgttgg tgtttcagta ttttttttt
3121 aaatctgtct gagggcaatg gctacttaat tctgtattcc tcctaaatct caaatttgca
3181 gcactttaag aattatggga ctactcttta tcttcttagc tgtcatattc tacttcattt
3241 aatttaaacc attttatcta cagctagttt atctacagtc acttcttcca aaattctaac
3301 tttttaaatt ttaattttaa ttattttatt ttatttttt gagatggagt ctcactctgt
3361 tgcccaggct ggagtgcagt ggcacaatct cggctcactg caacctctgc ctcctgggtt
3421 caagcgattc tcctgcctca gcctcctgag tagccaggac tataggcgcc tgccaacata
3481 cccagctaat ttttttttgt acttttagta gagattgggt ttcaccatgt tggccaggat
3541 ggtctcgatc tcctgacctc atgatccacc cacctcggcc tcccaaagtg ctgggattac
3601 aggtgtgaac caccgcgcct ggccttaatt tttatttta atttttgtgg gtacatagta
3661 ggtatatata tttgtgggat acatgagata ttttgataca ggcatgcaat gtgtagaaaa
3721 tcacatcatg taaaatgggg tatccatccc ctcaagaatt tatcctttgt gttacagaca
3781 atctaattat acttttagtt atttttaaat gtacaattaa attattgact acagtcaccc
3841 tgttgtacta tgaaatacta cgtcgtattc attctatttt tttttggtac ctgttagccc
3901 atcccctact acccttccca gcctccggta accattcttc cactctctat ctccatgcga
3961 ttgttttgaa ttttaaatcc cacaagtaag tgataacatg cgatgtttgt ctttctcaac
4021 attctaattt ttagtcattg caaacccatg ggggacagat tattcaataa gcttatttta
4081 ttttacttag cagccctaca gggcctatca tgggcactga tgcacctcac ctatttttac
4141 cttcatacca acgccctgag ttaaaaccat cctctcccca ttttgcagat gaggaagtga
4201 ggtgagctca cgtgcccagg gccgctcagc tggtatgtgg caggaccagg attccggccc
4261 acagcctgac tccagattcc tttcctttaa caaaatttta attttaatgt tatttaattt
4321 taatgtcaag ttatttttag gacattcgga cagactccaa aacaaatcag aagagatgtg
4381 acttttatgc aaactggaca ttagatgtgt tgataattaa ggagatagct ttaggccaag
4441 tggcctgcta cgggacctcc tgctggccag ttcctgcccc tatgttgggc caagtcacct
4501 ctccctcccc tccccttccc cctcctactc cagtcagtgt tcagggccca gggctcggca
4561 caggtgagac aagcaggact gaggggagag ggattcatgg gcgttcagag agaagagaga
4621 acagcagaat gaggagaggc ctaaagcctc c cgggggagc tggaggggaa ctgggaagga
4681 agagtggggc atcctaggaa gacgaacttg gccgagcaca gactcaggaa caagaaagcc
4741 gctgggagct gcccaggagg ccaaggggat tctcactcaa catccaacac cctacaagtt
4801 agaactgagt atctaaagtc tcctcttcca aggagcctgc ctgaacacac cagaaaacaa
4861 tgacccactt ccctcatctg acaccccacc cctcgcccgc ccctcggcca gtctcagctc
4921 ccatttctgc cccctatgtc tgtgcagggc ctgactgcct cgtcctcccc ctccttcctc
4981 atgtccacct tcctcccctc ttccccctcc tccccgttcc tgtatccttc cttcctcccc
5041 cacctcctcc cctttctctt cccagtctcc tccacctcct tcccccatac tcctcttcct
5101 ccccttcctt ccttctctcc tccccattct ccccctctcc cttttctgac cccttcttct
5161 ctcctcatca tcccctctcc tccctgctcc tccttcctcc tttcttccct tccctttcc
5221 tactcctctg cactcctcct cccctctttt ccagtcttct cctccaagcc ttctccttct
5281 cttccctctc ctcttgctcc ttcctctaac tctgtgctcc ctaaggggca gggcctgagc
5341 tcaggctcaa tgaaatccac tgaatgaaag caaacatggg gtggaatgct gcccctccca
5401 gctgagtcct tctctgacct tgggaagcag agggagaaag cagggaggga gctcacccca
5461 ttcccctccg gtccctggca gcctgggccc aggacacatg gtcctgtaaa tctccccagc
5521 ttataccatg ggctaaattt agacagaagc tttgtcgcca ggccctggag ggaacactag
5581 actgggccag acccccaggt tccagggagg cgataccttt gaggcctgtc ggggcccggg
5641 gcacaggttc tgtttggata ggacccaagg tggaagaaga gcagctggag tcttggctct
5701 cattcctgca tccttcccca gccagggcct ctcctttgcc tcttctaagg tgagggcctc

5761 ccatgccagg cactgagagg cgggaaccag gacctgggcc ctcacacagg ctaggctcag
5821 agcaaagttc agagagaagg aaatataaca agcaggaaaa gcccacaacc tggtctcccg
5881 ggattagggg tggggatgta gctctacctg gctgaggggg ccatggtttc attaatccat
5941 ccaagcagtc agtcagttga caaacattta ttgagcacct gctgtgtgcc aggctctgaa
6001 ctggggactg gggacacaga tgtggaccct gtgctcccac agcctacagt cagtctcatg
6061 ggctggggtg tgcatggtgg agagctctct gccctagaac agccaagagc taatcagcac
6121 aggtgtccaa gacgttcccc tcagaagaaa gaaaaatgga aacaggaaaa cattccattt
6181 cttctctaca ctcttacaca cctctgacac caaaagtgtg gattttctgc cccaagcaat
6241 tctccacttc tctggggacc ccaactgggt gtcctccaat tcaattcaat tctgatgcta
6301 tcttcctgga gttggcatca gcctccacag gtaaagcctc agtcccacaa gactgccccc
6361 aattcagatg ccagtcgcaa gacagggctt ctcattctcc caaccgacag gctatataaa
6421 tcaagagttc ccatgaccct ctctttgggt tcaatagtgt gctataacat tttgcaaaac
6481 tcaaggaaac atttctattt acaggtttat tttaaagtct atcacaaagg atatagctga
6541 acagccagat ggagaggcac acaggccaag ctgtggggag ggcgcagagc accgtgccct
6601 cactggatgc caccctcccc gcaccaccag gtggtcacca accagaagcc cccaatccca
6661 cctttgaggg gtttcacgga ggctccatta cagaggcata ttaacccaat ctccagcccg
6721 tctcccttcc ccaaggtgta ggaaggtgct gaaagctcca accttctaac ctgtagggca
6781 tttagaggca ggtctagcac tagagtaagt gctgacttca cgtgaaagaa gcctcccctt
6841 cctgcctgtc tcccctcttc ttctcttggc acctccatac cagatggtgc agtttcctgt
6901 gtgggccctg ggctcgcctg cctctgaggg tttgctttct cagcttctac ctgggtcact
6961 ttctcctctt cttactgctg ttcctatggc ccaagtcctt ccatcctttt agattccgct
7021 ccaatgtcca gttcgccgaa cggccttcgg tttcctcaat gagcattcac tgagcaccgc
7081 tgaggggcga gctcagtgag ggagatggag gctgcacctt caggggccca ttgtccgagg
7141 gaaagctgct gagggcaaac ggagtcccac actgactgga gagctgggga agggggggcac
7201 agggtttcca ggaagctttt ctgggcagga agagcttgag ctgggccttg ggagtcccga
7261 tgaggggact cgacctccca tcactgagct ctctaactag ccccctccta aagtccttgc
7321 caagagtcac cattcagcgt gcgcattcta cgtgccaggc tcttgaccta cagtaatttg
7381 ttgaatgcac acaaaatcct gtgggatttg tattatcccc attttaaaga tgaggagatc
7441 aaggctcaga gaacttaagt ggcttgccaa ggccacacag ctaagcccag catccccaga
7501 actctgcaca ctgggatgtg tgcactctac ccactctgac agctcttctc aagtccttgg
7561 aatgggatta tccaagatat gtctgtccct gactttcctg gaccatgtgt tttctgagga
7621 cagcagcctc tctttcctcc tcctccctgc ccacccatca ccaccaccac taggggctct
7681 gcccaaaatt gacacctgca tggtagatta cgacttcgga ctgagtcaaa tccacccaaa
7741 caagcccctt gaagagaatg agtatctttg gctgtgtcag cttttggggc cagcagagaa
7801 gttataggag atgaagccag gatctgggat tactccaaaa attacttctg atatgcctgt
7861 accaaatttt gcccagaccc aaagagcaag catcagttgc agagtatggt aaccctgtgc
7921 ccactcccca caaagtgctt gctcagccag aagatggcca cagcctgaga ccacagaatc
7981 caaggcaagg accatcccac attgcctgag tcctggggga ggatgccctg accccatcat
8041 gggttcaaag catgtgtcag aaaccccggt aaagtcaaac aggagatcac ctcagtgatg
8101 gggctacaaa tgtcacagct caggggcttt gagagaaggc actaacatct gagatggaat
8161 ggctgcatct gcctaagggc gcaccgtgga cactgcatgc ctctgggttg cccgtaccct
8221 gacctccaga ggctgggatt cactgggaac cctcccacag agggcattac tagctgcttc
8281 tcaaacaagg gcgatgggcg ccataaagag ggtgatgggt cagacctaga gccctcactg
8341 tccccagacc cagccaaaag tatggcctga gaattcttcc tggaaggtct cagtgtctgg
8401 agtgttccct ccctctccca gccccccaac ccctggcagt ggctcacggg gtcagcaagg
8461 tcttcctgat gcccgagaag ctggagctga ggaactctca gcagctccct tgatatgtgg
8521 acagcaagac catggccccc agtgtttgtc ctgcaggcca agaggccaga actcctggct
8581 gacagctcag gaggcccccа cagtgccgag acacacacac aacacacaca cctgctcccc
8641 tgcatttccc tctctacctc acttcgtggg gtgggagttt ggcagaacac aggtgcagtg
8701 agtgcctcca agtggtgtct gcctcaccct aggcttccgg gaggggtgca gggggcagca

8761 ctctcctgtc tgagcaggtg gggtcattca ttagaagctc caggctgacg tggtgagggc
8821 tgagcatgca gcttccagtc ccagcacagg aaatggactg agcccaatgt ggagaacgtg
8881 atgtttgagg gcttttgggc tctggcccct ggggacatcc gaattcttcc tggaaggtct
8941 cagtgtctgg agtgttccct cccctcccca gccccccaac ccctggcagt ggctcacggg
9001 gtcagcaagg tcttcctgat gcccgagaag ctggagctga ggaactctca gcagctccct
9061 tgatatgtgg acagcaagac catggccccc agtgtttgtc ctgcaggcca agaggccaga
9121 actcctggct gacagctcag gaggccccca cagtgccgag acacacacac aacacacaca
9181 cctgctcccc tgcatttccc tctctacctc acttcgtggg gtgggagttt ggcagaacac
9241 aggtgcagtg agtgcctcca agtggtgtct gcctcaccct aggcttccgg gaggggtgca
9301 gggggcagca ctctcctgtc tgagcaggtg gggtcattca ttagaagctc caggctgacg
9361 tggtgagggc tgagcatgca gcttccagtc ccagcacagg aaatggactg agcccaatgt
9421 ggagaacgtg atgtttgagg gcttttgggc tctggcccct ggggacatcc tggggcaaac
9481 attgacccag ctgccaggaa gagtacatct gacaggcagc aggaagagaa ggggctgcac
9541 actgatggag cagtgtcttc tggaatccca gaatcaggaa tctcagggtt gggaggggca
9601 caaaagtcac acacgtgtgt gcatgtactc acacacacac acgctgacca tagagcaaga
9661 aaaggagtgg tcttgggagg aatggggcca agcgttagaa tttagagtga attgctctcg
9721 ttcagtcata caccatgcat tttccagcct ctctgccttt gcacatgctg tgctctctaa
9781 ctagaatgcc cttcctggcc tgatcagctt ggcaagctcc tactcaccct tcaaaaccca
9841 cattagcatg acctcttctg tgatgctacc cctgccctcc tcagtaaatc gtacatgccg
9901 taaaactgta catcctatcc atctttccca ccagaagatg gtcccttcac gggtgcacat
9961 gcccacatct actttatccg tgcacatcca gcacccactc tgatgatttt ttaggagacg
10021 ggatcctggg tgcttttcta tcttttggat ggaattgcca ttattgcaac aatcacatat
10081 cttccctagt gagtgcctaa gctggcccta tggtcccttc ttttgtccct actcatcttt
10141 ctgggtgccc caaaatgcta cctctcccaa taagccttcc ctgtcttcct gaatccaggt
10201 taggcacttc tctcacgtat taacctaatg ctcttcgttg caagtgcctg cttacctgtg
10261 tgacagggaa acaaagttaa atcatgcatt tgcctcctct tccttctgca ctttctaagt
10321 ctagaggatc cctacctggc ctgggactac ctgaaggaca cccatcagtt cccccagcca
10381 gggcctgaca cacagtaggt gttctgcaaa tatttgttgg ttgaatgaac aactaatgga
10441 acccacccca cacccagctt tgctgacaga aggacacaag ccccatgaca gtgacacatc
10501 ccagcctgag gaggagacag ccttgggggc tgagggatgg acagagcaga cccctgccta
10561 ggaggcaaca ctgcattggg ccacttctga ggcctttctt cggtccagcc caggccagac
10621 tgctgctgct ccatgtggcc atgggcatct ccctccttgt tctccctctc tgtactaaga
10681 ggagtaaggg atgagtgata ggataggatg ggggcagaca ggcactgtca atggctagcc
10741 tgggcctgat ctttgaccct catccttgct gtcccctagg cagccgggca gagccagatg
10801 cccaaagacc agcccctgtg ggcacttctg gagcagtact gccagaccat catgaccctc
10861 accaacctct caggtaggta ccctctgcca ccaccctgca gcagcccctg gagatgtggg
10921 tgggccctgc gggaggaaga ggctttcctg gggccagccc aggctcagca gggtggtggg
10981 tggggtgctg gtttccagtc ctggtttgat ttctgccttg ctgagggccc ttacgcaggc
11041 ccttccccat tctaggctca ggcttgccac tgataatgaa gggcttggct gggaggccac
10101 tccccattcc cagctcccac gatggaagtc agctgggctg gcattcagga gtccaagaag
11161 gggatgaccc ctcagctttc caactccagg gaaactggca gaaggtggta aacattctct
11221 ggttgccgtg gctacaggtg tgacttcctg gtaaggactt gcaaagcagc agcccggccc
11281 tgagttctgg cccctcagag gtttccagtg gctgattcac agggtcaggg ccagccactg
11341 cagagctgca cagagccctg gattctgctg ggttccagag agatgtctcc agtgtcccac
11401 cctgtatcct ggccttccca ggtctgcatc cacaggatgg gagtgggggt gacccttgct
11461 cgggcctggg gccactgcca gggcacaagg agacaagggt gctggccagc acccagaccc
11521 agcctacagg cactgcccaa gccaggcgtc aaagagaaag tcttgctgcc tgttctcctt
11581 ccactggggc agtcaatagc tccggtgttg ggatcaggta gatccaagtc tggatcctag
11641 ctcttccacc cagcggcctc gcccagcccc tggaaactca cctaacccat ctgtattagt
11701 gtcctggggc cactggaaca aagcaccaca gactgggggc ttgagcagaa gctgatttca

11761 tcaccattct ggaggctggt aggccaagat caagctgtgg gcagggttgg ttcctcctga
11821 ggcctcccta cttggcttgc agatggccat ctttctccct atgtcttcat atggtctctc
11881 ctctgcatgt ctgtgtccca atctcttctt cttctaagaa caccagtcct attggattaa
11941 agcccgctta ccttacttta ctttaatcac ctcatcgaag accctaaatc caaacaaggc
12001 cacattctga ggtactggga gttaggactt caacatacaa attttggagc aacacagttc
12061 agcctgtgac accatctgaa gctagtttcc tcatctgttt agcaatggat ctaagtccac
12121 ctcggaggct cacagtgaag gctgaatga gagcccacggg gacacctagc acaggccatg
12181 ccagcatcag acaggaaatg gaggcccctc ccttcaggaa agcccaacct tccctcaccc
12241 tcagaggcag cgtgggtttg tcagcagacc catagccttc tccctgcctt cctctgccct
12301 ccctccagca ggtggcacag acccttccct gtctcccttg gcggatggtc cccgctctga
12361 gtgcagcgga tagggtccca gcttctccaa acatacagcc ttgcctgttc agcttctaac
12421 agcattcatc catttgcaaa agggatccac atgatttcta taaaattggg cacaaaaaaa
12481 tttaactcaa aggagaagat gataagtact cattctccca caacccagga atttttataa
12541 tgcctcatat ccctcaagaa tttatgcatt catagagcct catgttcaat tgttttttta
12601 ttttcttttt cttttctttt ttttttaatg ggataatagt gtgtccatag tttcgtcgcc
12661 tgcttttttcc tcttataact cagagtgaac atttccccat gtcattaaat attcaaccac
12721 cgcatcgctt gtgacgctgc ctagcagcca ctctggggat gcatcacccc acatgaacca
12781 tgaccggatg gtggatgtag aggctgctag ggagcccctc gagcttcagg aatcatttct
12841 caagtcccca aattgataat gaccttcttt tcagtgtcca ctgcttgggg aaatgcatca
12901 tgtcta acgc agagttcaaa accaagtaaa accaaacact atatctttta gatgtacata
12961 cctgtgagaa aactgaaaaa aaatatgagc cagagaatga caaatgcaat agtcaaggga
13021 atgttccctc tggggaagga gaatagaaat aaatagataa gtgactggta atgatgtagt
13081 ttccaattgg gtggtaggctt attatattg tctactacat acaataaaaa gactaaataa
13141 aataaagact aaaaaataaa gttaagccat gcatgggcta atgcacagac taataatgac
13201 aaagggtcac gaatctatga tcaattcaat tgtatgcccc tgaaaaaggc aacgcatgca
13261 aaatcacaga gcaattatga cttcaggcag ggtttaaata gattatttct attttattca
13321 tcacagaggc agctacatat atttgaaaaa tattagggca tatattatga ggcatattat
13381 ttatcctatc catagatgat ataaaaatcc aaatggacta tttatttatt ttttattatt
13441 ttttgtaggc tggtcttgaa ttcctgtgct caagcaatct gcctaccttg gcttccccaa
13501 gtgctgagat tatgggtgtg agccactgca cctggccaag aatccgaatg gattcaaaga
13561 taccttgaaa taattcctca atgcaacaca cacacatatg ccagggttgg tcaaatggga
13621 agagagcctt ggcctgaaga cagggacctg ggctttcctc agctcttctg ccaaggtatc
13681 tgtcctttct tagtgactca ctgggctgaa gtctagagca ttccaatggg tgctggggat
13741 gggttagtga accaggacca gccctgccct tggggaggca tgggctggta gatgagacaa
13801 tgaataaaaa gcaacctgga tgtgacatag gccagcaccc agcagatggg gtcaccaagg
13861 agctgcatgt ctgaaggatg agtgtggagc tgtaaggcca tttccagtgc agaaatacaa
13921 acaaggaaca gagatagcga atggtttctg gtagtgacac attgttctgg agggcctttc
13981 aaatgaagca ggagggtgag gatgggggac gatgatcatg aagaaccttc tttgcaatac
14041 caaagatggt tcccaggaat gacatgttgt tctggagggc ctttcaaatg gagcagaagg
14101 gtgaggatga gggatagatg attacgaagg accttctttg caataccaaa gaggtggtct
14161 ccatctcaag ggcaatgggt tggcggcact cctatgagga agagtgacag gatcgtaatt
14221 ataaaacatc gttttatctg ccatgaggag acagggagcc catttgggct gctgcagaga
14281 tgaggatggc ctgggagaag taagggaggg aaacctcttt tgcttgtctg ggaatcctat
14341 tgtccatctg taaaacaaga atgtgggatt ggatgtcact gagagtcctt tctgtcccca
14401 cagcttagtc ccatggtttg acatgaagac accaggccaa ttcctctgtc ctgcagtgga
14461 gtaggatagg gtgaggaggg ggtgcctggc ctgggacccc tgtctccagt gctacagggc
14521 aaaagtgcag caccttcagg cccctctgga acctctgtgc cctcagcaac cccgtcgtaa
14581 ttcccaggaa acaccacgac tattccagct gtgcctgaat agactccctc tctatgtagt
14641 ctaaccaagg acagtccctt agagaagcaa agatgcatcc tgtcccttta aatcctgttt
14701 tccagctgac atttgagtgg taggggatga attagagaga gaatgtgtgt gcatggtgaa

14761 gatgcataac tgtggctctg tgtgtgattc tgttgtgcct acatgcctgt ttgatgccat
14821 gtattaggct ttgcaagtgt gtctgagctg gcctggaact gtccactgct gctgcagctg
14881 acatcagagg tgggccatgg gatgggatgc agggcaccag aggcacctgc cttactctgc
14941 tgctcatggt acacaggggt cttccaaagt acctgtcaca cttccctgaa cctatttgcc
15001 aacctgtccc caacagcttg gggacacaca aactgttcca aatacttatc ccctaattcc
15061 tggcctccag ctgggatggg gctggcctgc agcgctggga acccatcact atcccaaagc
15121 ctctaatcta cctctgcttc tttagttagc aaaaatgccc ctgtctttgt ttgtttgtac
15181 ttggatttag taaaattgag ggaattttgg ggctcttccc aatttcttcc tcatttgcct
15241 gtttgcagac a taaagtga gctgtaaaat caatttgttc ccaaactgct acctttttcta
15301 gttttccctc tgtcacatct cgaatgacaa ctgtgcctat aataataaga tccatgaaga
15361 aatggcccca catccagggg acctcgggtc tgtgggtctg tgggtatgtg ctcagacccc
15421 aactgctcat tcaggatgca gaacagcctt tgactctgcc actgaaatgg tcacttccca
15481 gaagaatctt tggtatgtga gctctcccag ttaagagcaa tccagttcac ccaggccaac
15541 cctctgttct gcagaaatct gcatgagatg agaagtccct ggcctctggt cattgggagc
15601 acaccacctc ttgacatagc tgcatctata ggtgctttag gcaccaccta ggtggtggct
15681 gcagtaatga acaaggataa taacttaagc ttggcttagg atttctccct tcctctgtgc
15741 tgcaccgtgc tgtgtgaagg gctatccctt gtatatgagg gacactattg ccctgtgaat
15801 acatgggctg acctggccag ccaccctctg gcctgtgtgg gaccctgggt aacagggctc
15861 agtctaggag cggagggtag gactctgcct aggtcctccc ctgctcctgg cactggataa
15921 tatgaagacc aaggaacctc tacctctgca gctgcccaga gctgagcctg ggctctcact
15981 gtccattcag gaggcagagg gtagacctgg ctggctgctc agcatcttgt tcctgccagt
16041 cagcaggccc tggagagatt ccagggctca gccctggtct tagtgtggtc tcactcgctc
16101 tgcctgggca gtgacagcac cattaatatg gaggctggtg agagcggagc acacaaaagc
16161 agcctgcctg ctgctttgcc tctctctgcc cagggcatgg tgctagttca tggtggtttc
16221 agcctttcct agcagcttag gtttatgtgg aggatggcag gggaacagga ctggtttctg
16281 agactaggtt ccagctctcc ttgccctcta aagatagaaa caaacaacac acaacacaca
16341 tgtatgtctt ccccagactc tccgtctcat aactcagaac cagagagtct cagagctgag
16401 aggggcctcg gggaagattt cactgatgga gaaagctcca gaagagaggg caagcatcct
16461 gcctggggta ccatggcagg gagagcagag tcagaggctg gaccaaaatg cagcctgagt
16521 ggaagcatgc cttaccccag gccctgcttc ccgagcccaa tgccccactc agctggcaga
16581 cacagcaggc ccgccctggg cactctgagg gtgcagccac agcagatcac caaggaggca
16641 ttctgggcca gggtggggtg gggggcctgg gcaagttctc tggggaggtt tccagctcct
16701 ccacacctgc tgtggggcct gattctcccc gcccctgccc cgctactggt gtggaaacca
16761 gggtcaggtg aagccctgcc caagccttga ggaagagaga cacattctca cttcttcctt
16821 tattctttct ggtgcccaag gcactaacaa ctgggtgtat aaagccttcc agataatttc
16881 agccaatttc tcccctagac ttactccatc tgattaaatg gccacccagt cactcaagca
16941 ggagacctga tgttatccct gcctctgggc tcctacttaa aaccagccac ccactccagt
17001 tcctcctcac caggtctggc tgcccctccg aaatctctct ccattccagc ttccacggtc
17061 tgatccaagg acttcaaccc actcaggcca cttgatctca ttagactcat gggtgtctcc
17121 ctgcttccct ccaggcctcc tcgatacatt tccttcccag aggctatgga ggcttttgga
17181 ggatgcagat ctgcccatgg gctctccgtc tctgcctccc tccagtgtcc tcgctcaggg
17241 aggggaagct caggcggaaa gctgccgaac tttgggttgc ggctgtccct cgattagcag
17301 agctgcggtg ttctcctcgg gcaggcgggc aggtgggcgc gctttgctgc cccctgcagc
17361 tcgggggcct gcgatccccg cacagagcat tccgtcaccc caggcccacg ctctccagcc
17421 caccgccctc ctctggacgc cgcgagtgga agagagctgc gaactgagaa gccgtacttt
17481 gggcagggtg gagggcccgg gggctggaga ctgagcccct ccgagaggag ccgcccggcc
17541 ccgccccccg gcgcagccat tggccgcggc ggagcggctg tacccgcagc tccgtgcact
17601 cggcggctcc tctccgggaa ggtccccact tgacagctct gggcgaccgg aggtggcgcc
17661 caaaggctgc ccgggagatc ggggctgggc tggcgggggc caggaccccg cgccctctcg
17721 gccgctcact ctcgcgtcca ctccctcgca gtcacgccgg gcgcgcactc ccactccctc

17781 tccgcacgcg gctgcgggac gcgatggacg cggcactgct ccacagcctg ctggaggcca
17841 actgcagcct ggcgctggct gaagagctgc tcttggacgg ctggggccac ccctggaccc
17901 cgagggtagc ggtcgaagct gcgcgggctc cttctctcgc gcgctgggtg cgcccctgac
17961 cccgcgctcc gagagccccg cgtccctgcg ggcccctctg agccctactg cgctctctcc
18021 cgccgccagg tccctactcc tactgcaaca cgaccttgga ccagatcgga acgtgctggc
18081 cccgcagcgc tgccggagcc ctcgtggaga ggccgtgccc cgagtacttc aacggcgtca
18141 agtacaacac gaccc